Supplementary Material

The Supplementary Information contains Supplementary Materials and Methods, Supplementary Results and Discussion, Supplementary References, Supplementary Figures S1 to S12 and Supplementary Tables S1 to S6.

Supplementary Materials and Methods

Environmental measurements and sample collection

Samples were collected in sterile 50 mL tubes in June 2010 at three sampling sites (B1A: up to 2 cm deep; B1B: from 2 to 15 cm deep; B2: 50 cm deep) determined by the presence of each different macroscopic microbial growth morphology and kept on ice until nucleic acid and protein extraction was performed (within the following 2 h). Conductivity, temperature (T), Eh, pH and dissolved oxygen (DO) concentration measurements were conducted at the sampling points using a Multi 340i multi-parameter meter (WTW, Germany) from June 2010 to June 2012. Conductivity calibration was carried out with a KCl 0.01 mol/L control solution. Reference solutions with pH values of 2.0, 4.0 and 7.0 were employed for pH meter calibration. The DO probe calibration was performed in water vapor-saturated air. All geochemical data were imported into the R statistical computing environment, version 2.15.0 (Oksanen *et al*., 2010), and all subsequent analyses and visualisations were performed using functions of the ggplot2 packages (Wickham, 2009), version 0.8.9. Al, As, Hg and Pb concentrations were determined through Inductively Coupled Plasma-Mass Spectrometry (ICP/MS), whereas and Fe (II) and Fe_{total} contents were measured using the modified ferrozine method, as described elsewhere (Viollier *et al*., 1999). XPS analysis has been performed with a Fisons MT500 spectrometer coupled to a hemispheric electron analyzer (CLAM2) and a X-ray source which utilizes the Kα radiation of Mg (1253.6 eV) and operates at 300W. B1A, B1B and B2 dried biofilms were homogenized into powder and set in the

samples´s carrier by means of a double-sided conductive tape to be introduced into the spectrophotometer. Residual pression in the analysis chamber was maintained under 10-8 Torr during the measurement. Spectra were registered using an energy of 20 eV. Component analysis for each element was carried out by a subtractive analysis of the background based on the Shirley method and by adjusting the experimental curve to a mixture of Gaussian and Lorentzian lines in a variable proportion. As charge effects due to low conductivity levels were expected, binding energy of C 1s, derived from saturated hydrocarbons at 285.0 eV, was utilized as reference in the calibration of each spectrum.

Nucleic acid extraction and full-length 16S rRNA gene clone library construction

Nucleic acid extraction was performed directly from cooled samples using the PowerSoil® DNA Isolation Kit (MoBio, USA) according to supplier's recommendations. Sequences of 16S rRNA genes were amplified via PCR using AmpliTaq Gold® 360 Master Mix (Applied Biosystems®, Life Technologies™, USA) following the manufacturer's instructions. The following primer pairs were used at a final concentration of 0.4 µM: 27f/1492r, for *Bacteria* (Hongoh *et al.*, 2003; Lane *et al.*, 1991); 20f/958r (Massana *et al.*, 1997; DeLong *et al.*, 1992), for the *Archaea* domain; and ARM979F/ARM1365R (Baker *et al.*, 2006), for "ARMAN"-like cells. PCR amplifications were performed in a MJ Mini Thermal Cycler (Bio-Rad, USA) with an annealing temperature of 55ºC for 30 cycles. The PCR products obtained from three separate reactions for each primer set were pooled and purified using Illustra GFX™ PCR DNA and Gel Band Purification Columns (GE Healthcare, UK) and quantified on 1% agarose gels using the Low DNA Mass™ Ladder (Applied Biosystems®) as a reference. The 16S rRNA gene sequences were cloned using pGEM®-T Easy Vector System I (Promega, USA) according to the manufacturer's directions. After plating, the positive transformants were screened via PCR amplification of the inserts with flanking vector primers

(M13F 5´-GTTTTCCCAGTCACGAC-3´; M13R 5´-GAAACAGCTATGACCATG-3´), and amplicons were sequenced according to the protocol of the BigDye Terminator v3.1 sequencing kit (Applied Biosystems®) and subjected to capillary electrophoresis in an ABI PRISM® 3130xl Genetic Analyzer (Applied Biosystems®).

SSU rRNA hypervariable tag sequencing

V1-V3 amplicon libraries were constructed using the primers 27f/338r for the *Bacteria* domain, 344f/519r for the *Archaea* domain and ARM979F/ARM1365R for "ARMAN"-like cells (Baker *et al*., 2006). Primers were bonded to ten-nucleotide multiplexing tags and to the 454 FLX sequencing adaptors A (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3') and B (5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3') in the forward and reverse regions, respectively. Amplification of the V1-V3 regions was performed in 25 µL reactions using AmpliTaq Gold® 360 Master Mix (Applied Biosystems, USA) following the instructions provided by the manufacturer, with primers at a final concentration of 0.4 µM. Each amplification began with a denaturation step of 95ºC for 10 min, followed by 30 cycles of 95ºC for 45 s, 55ºC or 60ºC (as the annealing temperatures for 27f/338r and 44f/519r, respectively) for 45 s and 72ºC for 1 min, with a final extension cycle of 72ºC for 10 min. The products obtained from six separate reactions were pooled and purified using Gel Band Purification Columns (GE Healthcare, UK), and the DNA concentration was determined using the Quant-iT™ PicoGreen® dsDNA assay kit (Applied Biosystems®, USA). The uniquely tagged, pooled DNA samples were immobilised on DNA capture beads, amplified through emulsion-based clonal amplification (emPCR) and sequenced in a 454 Life Sciences Genome Sequencer FLX (Roche Diagnostics, USA) using 1/16 of a plate. A total of 214,427 raw bacterial sequences obtained from 454 pyro-sequencing were processed using the QIIME pipeline (Caporaso *et al*., 2010). Reads with lengths below 200 bp and quality scores of less

than 25 were excluded from further analysis. No mismatches were allowed in the forward primer. Using the default settings, 23,751 quality-controlled sequences were denoised and binned into 1,005 operational taxonomic units (OTUs) at a 97% similarity cut-off (Kunin *et al.*, 2010) using uclust 1.2.22 as an OTU picking method (Edgar, 2010). The cluster seed was employed as a representative sequence. The Chimera Slayer algorithm detected 7,565 chimeric sequences, which were excluded from subsequent analysis. Following exclusion, 6,346 sequences for sample B1A, 9,000 for sample B1B and 8,405 for B2 were aligned with PyNAST (Caporaso *et al*., 2009) using the Greengenes core set alignment as a reference (DeSantis *et al*., 2006), and taxonomy assignations were inferred through comparison with both the Ribosomal Database Project (Cole *et al*., 2009) and BLASTn (Altschul *et al*., 1990) databases. The alignments were filtered to differentiate among conserved and uninformative positions for tree building. An attempt at phylogeny reconstruction was made using FastTree 1.2 open-source software (Price *et al*., 2010), although these data were not taken into consideration here due to the lack of accuracy of the phylogenies constructed from 454 reads. The sequences were rarefied to remove the heterogeneity in the number of sequences per sample prior to the calculation of alpha and beta diversity statistics. Alpha diversity was measured through the Dominance, Simpson, Shannon, Evenness, Brillouin, Menhinick, Margalef, Equitatibility, Fisher-alpha, Berger-Parker and Chao-1 indices. Beta diversity was calculated using weighted and unweighted UniFrac metrics and PCoA plots, visualised with KiNG display software (Chen *et al*., 2009). For *Archaea* (including "ARMAN"-related sequences), a total of 74,124 raw sequences were included in an analysis using QIIME software. In this case, a minimum sequence length of 150 bp was used; 3 mismatches were permitted in the forward primer, and sequences with quality scores below 20 were discarded. Following quality control, 52,428 sequences were included for further analyses. The final number of non-"ARMAN"-related archaeal sequences per sample following chimera

detection was 3,496 for B1A, 14,727 for B1B and 2,495 for B2. For "ARMAN"-related sequences, there were 10,333 sequences obtained for B1A, 9,524 for B1B and 11,853 for B2. A total of 637 archaeal OTUs were detected (383 from non-"ARMAN"-related and 254 from "ARMAN"-related *Archaea*). Successive analyses were carried out following the same steps used for *Bacteria*.

Phylogenetic analysis

The obtained Sanger electropherograms were corrected and checked for the presence of chimeras against the Greengenes database (DeSantis *et al*., 2006) using bellerophon (Huber *et al*., 2004). Full-length 16S and SSU rRNA sequences were automatically aligned against the SILVA SSURef_111 (Pruesse *et al*., 2012) and LTP_s108 (Yarza *et al*., 2008) reference alignments with the SINA aligner (Pruesse *et al*., 2012) and were manually examined to correct inaccurate bases using the ARB package (http://www.arb-home.de/). Maximum likelihood phylogenies were constructed using general time-reversible distances to define the nucleotide substitution model. For the phylogeny of the "ARMAN"-like sequences, a specific SAI filter that included 977 characters was employed. To verify the absence of phylogenetic incongruences, the resulting tree topologies were tested against the currently accepted classification of prokaryotes (LPSN, http://www.bactreio.cict.fr). PAST software, v2.17b, was used to calculate the microbial diversity indexes, including dominance, Simpson's index, Buzas and Gibson's evenness, Brillouin's index, Menhinick´s richness index, Margalef´s richness index, Equitability, Fisher's alpha, Berger-Parker dominance and Chao-1.

Streamer formation dynamics assay

To investigate the streamer formation dynamics, 300 coat-cleaned slides were submerged at B1 acid mine drainage stream in a holder that maintained them vertical, in a region of early

streamer stages. Collection of slides was made at different colonization phases: 6 hours (80 slides), 1 day (80 slides), 2 days (40 slides), 4 days (40 slides), 10 days (20 slides), 23 days (20 slides) and 58 days (20 slides). Samples for FISH analysis (including 2 slides from the 1 h stage) were fixed *in situ* with formaldehyde diluted up to 4% in acid water (pH 2), kept at 4°C for 6h, washed with PBS until complete removal of formaldehyde and stored at -20°C in ethanol : PBS (50% v/v) prior to hybridization. The slides were scraped with sterilized sticks and DNA extraction was performed with the PowerSoil® DNA Isolation Kit (MoBio, USA). V1-V3 amplicon libraries were constructed using the primers 27f/338r (Hongoh *et al.*, 2003; Fierer *et al*., 2008) for the *Bacteria* domain and 344f/519r (Raskin *et al*., 1994; Ovreås *et al*., 1997) for the *Archaea* domain. Primers were bonded to ten-nucleotide multiplexing tags and to the 454 FLX sequencing adaptors A (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3') and B (5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3') in the forward and reverse regions respectively. Amplification of the V1-V3 regions was performed in 25 µL reactions using the AmpliTaq Gold® 360 Master Mix (Applied Biosystems, USA) following the instructions provided by the manufacturer and using the primers at 0.4 µM final concentration. Each amplification began with a denaturation step of 95°C for 10 min followed by 30 cycles of 95°C for 45 s, 55°C and 60°C as annealing temperature to the pairs 27f/338r and 344f/519r respectively for 45 s and 72°C for 1 min. The final extension cycle coursed at 72°C for 10 min. Products from 6 separated reactions were pooled, purified using the Gel Band Purification Columns (GE Healthcare, UK) and checked for concentration with the Quant-iT™ PicoGreen® dsDNA assay kit (Applied Biosystems®, USA). The uniquely tagged, pooled DNA samples were immobilized onto DNA capture beads, amplified through emulsion-based clonal amplification (emPCR), and sequenced on a 454 Life Sciences Genome Sequencer FLX (Roche Diagnostics, USA) using 1/16 plate. A total of 17,824 raw sequences from 454 pyro-sequencing were processed using the QIIME pipeline (Caporaso *et*

al., 2010) as described for mature communities B1A, B1B and B2. A total of 15,006 quality-controlled sequences were denoised using default settings and binned into 614 operational taxonomic units (OTUs) at a 97% similarity cutoff. Alpha diversity was measured through Dominance, Simpson, Shannon, Evenness, Brillouin, Menhinick, Margalef, Equitatibility, Fisher-alpha, Berger-Parker and Chao-1 indices.

DNA sequencing, assembly, gene prediction and annotation

Sequencing was performed with a Roche 454 GS FLX Ti sequencer (454 Life Sciences, Branford, CT, USA) at Lifesequencing S.L. (Valencia, Spain), with one picotiterplate. Assembly was performed with Roche Newbler assembler v. 2.5.3 using the default parameters. An estimated error rate (incorrect bases/total number of expected nucleotides) of 0.49% was considered for the GS20 reads (Huse *et al*., 2007). The error rates for the GS20 reads were calculated using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970). Sequence information as follows:

B1A

Number of reads: 487,368 sequences with a mean read length of 335.34 nts that provides a total number of 163.43 Mb. These sequences were assembled using Newbler version 2.3 with the parameters by default, and provide a total amount of 68.8 Mb of estimated metagenome (2.4 x coverage). Approximately 55.25% of the sequences were introduce in the assembly, providing a total amount of 23,721 contigs, where 13,760 contains were higher than 500 nts and the N_{50} 1,054 and the largest contig 21,838 nts (for contig length distribution see **Fig. S12**).

B1B

Number of reads: 454,404 sequences with a mean read length of 391.15 nts that provides a total number of 177.89 Mb. These sequences were assembled using Newbler version 2.3 with

176 the parameters by default, and provide a total amount of 60.3 Mb of estimated metagenome (2.9 x coverage). Approximately 63.85% of the sequences were introduce in the assembly, 178 providing a total amount of 19,405 contigs, where 12,020 contains were higher than 500 nts and the N_{50} 1,506 and the largest contig 42,630 nts (for contig length distribution see **Fig. 180 S12**).

B2

182 Number of reads: 302,143 sequences with a mean read length of 374.97 nts that provides a total number of 113.29 Mb. These sequences were assembled using Newbler version 2.3 with 184 the parameters by default, and provide a total amount of 47.6 Mb of estimated metagenome (2.4 x coverage). Approximately 60.55% of the sequences were introduce in the assembly, 186 providing a total amount of 16,143 contigs, where 9,869 contains were higher than 500 nts and the N_{50} 1,126 and the largest contig 15,206 nts (for contig length distribution see **Fig. 188 S12**).

In order to create a relative trustworthy data set, we first disregarded contig sequences < 190 200 bp. Secondly, we removed all predicted ORFs < 50 bp, so the shortest ORFs are ~75 bp long. Potential protein-coding genes were identified with MetaGene (Noguchi *et al*., 2006). 192 Additionally, contigs without open reading frame (ORF) predictions by MetaGene were translated into artificial reading frames and were BLAST-searched against the NCBI-nr 194 database for similar sequences. Artificial translations with similarities were further processed in the same manner as the predicted ORFs from MetaGene. Annotation was performed with 196 GenDB, version 2.2 (Meyer *et al*., 2003), supplemented with the tool JCoast, version 1.7 (Richter *et al*., 2008). For each predicted ORF, observations were collected from similarity 198 searches against the NCBI-nr, Swiss-Prot, KEGG and genomesDB sequence databases (Richter *et al*., 2008) and against protein family databases from Pfam (Bateman *et al*., 2004) 200 and InterPro (Mulder *et al*., 2005). SignalP was used for signal peptide predictions (Bendtsen

et al., 2004) and TMHMM was used for transmembrane helix analysis (Krogh *et al*., 2001). Predicted protein coding sequences were automatically annotated using the in-house software MicHanThi (Quast, 2006). The MicHanThi software predicts gene functions based on similarity searches using the NCBI-nr (including Swiss-Prot) and InterPro databases. The annotation of proteins highlighted within the scope of this study was the subject of manual inspection. For all observations regarding putative protein functions, an E-value cutoff of 10^{-05} was used. To identify potential metabolic pathways, genes were searched for similarity against the KEGG database (Kanehisa *et al*., 2008). A match was considered valid if the similarity search resulted in an expectation E-value less than $1e^{-05}$. All occurring KO (KEGG Orthology) numbers were mapped against KEGG pathway functional hierarchies and were statistically analyzed. All predicted ORFs were also examined for similarity against the COG database (Tatusov *et al*., 2001). A match was considered valid if the similarity search resulted in an E-value less than $1e^{-05}$. JCoast's internal Last Common Ancestor (LCA) algorithm has been used for taxonomic affiliation of contigs.

Richmond mine AMD metagenome for comparative genomics was obtained by the IMG/M webpage of the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/) (Markowitz *et al*., 2012). For each metagenome (including those from *Los Rueldos* and *Richmond*), the functional classification by KEGG has been downloaded. All of the downloaded lists were parsed by counting matches to KO entries in the original KEGG hierarchy. The final list was normalized by calculating the percentages of occurrence. This list was used to build heat maps with JHeatChart (http://www.javaheatmap.com). Additionally, a distance matrix was calculated between all of the metagenome KEGG pathways listed by measuring the dependence between the two lists with the Pearson product-moment correlation coefficient. This distance matrix was used to generate the tree and was finally integrated into the heat map.

Note: the mine drainage metagenome B2 (as referred in the study) has been deposited in NCBI database as B2A metagenome, as a three letter code is need for submission under the umbrella BioProject.

Metaproteomic setup: protein extraction, liquid chromatography, tandem mass spectrometry and protein identification and quantification

Whole-cell protein extraction was performed by heating a (v/v) mixture of disruption buffer and microbial aggregate at 80ºC for 1 h, including a 2 min ultrasound bath every 10 min. The disruption buffer contained 150 mM NaCl, 2% sodium dodecyl sulfate (SDS), 100 mM ethylenediaminetatraacetic acid (EDTA), 1M Tris, 100 mM 1,4-dithio-D-threitol (DTT) and a ¼ tablet of Complete protease inhibitors (Roche Applied Science, Germany) to 1 mL of buffer. This step was followed by 7 cycles of 5 s sonication, 5 min centrifugation at 15,000 x g and spin filtering for 7 hours of the supernatant using Vivaspin filters (Sartorius, Germany) with a molecular weight cut off of 10,000 Da, after SDS dilution using 20 mM triethylammonium bicarbonate buffer (TEAB). Urea was added prior to spin filtering until 4 M final concentration to accomplish a better recovery of proteins.

Quantitation of the extracted protein amount was carried out by means of the Bradford Protein Assay (Bio-Rad) (Bradford, 1976) and a total protein amount of 60 µg per sample (divided over 3 gel lanes) was used for further analysis. Each gel lane was divided in five slices aiming to roughly get equal protein amounts. Gel pieces were sheared into ~1 mm^3 cubes and washed twice with water/acetonitrile 1:1 (v/v) for 15 min. After liquid removal, acetonitrile was added until the gel particles had shrunk. After removal of the acetonitrile, reduction was performed with 10 mM DTT in 0.1 M TEAB at 56ºC for 45 min, followed by alkylation with 55 mM iodoacetamide in 0.1M TEAB at room temperature for 30 min in the dark. Gel particles were rehydrated at 4ºC using a digestion buffer containing 5 ng/µL trypsin

in 20 mM TEAB with 1% sodium deoxycholate (SDC). Digestion buffer was added to completely cover the gel cubes and after 45 min, the remaining supernatant was removed and replaced with 5-20 µL of the same buffer without trypsin in order to keep the gel particles wet during enzymatic cleavage, at 37ºC overnight.

After digestion, peptides were extracted from the gel pieces by incubating them twice in a 0.1% SDC aqueous solution for 30 min at room temperature, while shaking. The same procedure was repeated using acetonitrile. Extracts collected at each step were pooled and SDC removal was achieved by phase separation with ethylacetate after acidification with trifluoroacetic acid (TFA) as described (Masuda *et al*., 2008).

Equal aliquots of peptides per sample/location extracted from the five gel slices were analyzed by nano LC-MS/MS. Peptide extracts were diluted using a 1% (v/v) TFA, 10% dimethyl sulfoxide (DMSO) (v/v) buffer. To afterwards enable the estimation of protein abundances, also the following internal standards were added to each sample prior to injection: 100 fmol of digested bovine serum albumin (Waters Corporation, USA) and 50 fmol of digested rabbit glycogen phosphorylase B (Waters). Peptide separation was performed with a nanoAcquity UPLC system (Waters) fitted with a 2cm, 180 µm ID Symmetry C18 trap column (Waters) and a 25 cm, 75 µM ID, BEH C18 (1.7 µM particles) analytical column (Waters). Peptides were trapped for 2 min at 10 µl/min with 0.1% TFA and separated at 350 nL/min using a 75 min gradient of 5-35% acetonitrile with 0.1% formic acid. The nanoLC was coupled to an LTQ Orbitrap instrument (Thermo Scientific, Germany) operating in data-dependent acquisition mode to automatically select the 10 most intense precursor ions for fragmentation by collision-induced dissociation (CID). Survey MS-scans were acquired from m/z 300 to m/z 1650 with a resolution of 60,000 (at m/z 400). MS/MS spectra were acquired in the linear ion trap (target value of 10,000 ions) with a maximum fill time of 100 ms. Lockmass recalibration was disabled as an active background ion reduction

276 device (ABIRD) (ESI Source Solutions, USA,) was used to reduce the level of background ions.

278 Proteome Discoverer v1.3.339 (Thermo Scientific) was used to process the raw data files and perform database searches on an in-house Mascot server, version 2.3.02 (Matrix Science, 280 UK). The data were searched against a combined 'common' database (B1A, B1B, B2 combined, annotated and made non-redundant). This database contained a total number of 282 49,389 unique/non-redundant sequence entries. The database used for Mascot searching contains a total of 49,432 unique sequence entries (including the standards and common 284 contaminants). Trypsin was selected as proteolytic enzyme, allowing a maximum of 2 missed cleavages. Search tolerances were set to 8 ppm for peptide precursors and 0.6 Da for fragment 286 ion masses. Carbamidomethyl cysteine was specified as fixed modification whereas oxidation (M), deamidation (N/Q) and N-terminal protein acetylation were set as variable 288 modifications. The Percolator tool was used for peptide validation, followed by high confidence (q-value ≤ 0.01) filtering which applies the correct cutoff value to achieve a $\leq 1\%$ 290 false discovery rate (FDR) on peptide level. Proteins were only included if they possessed at least one unique peptide scoring above the calculated cutoff value corresponding to the 292 desired FDR.

The combined 'common' (B1A, B1B, B2 combined, annotated and made non-redundant) 294 database was appended with common laboratory contaminant proteins as well as with the two internal protein standards to address technical variation and to enable label-free absolute 296 quantitation (Silva *et al.*, 2006). The approach is based on the best ionizing (most abundant) peptides per protein and allows estimation of absolute protein abundances within a sample 298 (stoichiometry) to assess dynamic range as well as to investigate differences between the acid mine regions. Protein area values were generated in Proteome Discoverer and exported to 300 Excel. The known molar amount of internal standard was used to calculate the absolute

amount of each protein on-column. Protein abundances per sample location were then
302 expressed as mmol/mol by dividing the individual protein amounts by the sum.

304 *Metaproteome-scale metabolic reconstruction*

Available network reconstructions in the literature are organism specific. Genome-scale
306 metabolic networks provide a more accurate description of metabolic processes of specific organisms, typically including physiological information manually curated, as well as in-
308 silico algorithmic refinements. Aside from reactions and their associated information (reversibility, compartment, EC number, genes), they may involve additional experimental
310 information. Accordingly, metabolic reconstruction typically starts from an annotated genome, which is then integrated with reaction databases and reported physiological
312 evidences, as well as with gene transcript levels or protein expression values to produce a functional computational model.

314 In our case, the reconstruction is slightly different as we aim to reconstruct the entire metabolic network where multiple proteins from multiple organisms are co-expressed as it
316 was recently reported (Hernández *et al.*, 2013). Briefly, herein we used the expression values of a set of proteins that were unambiguously quantified. We used the Enzyme Commission
318 number (EC number) as a numerical classification scheme for enzymes based on the metabolic reactions they catalyse, with each EC class exclusively defining the function
320 performed by the enzyme (Bairoch *et al.*, 2000). An EC number in this analysis is used exclusively to refer to an enzyme. Using appropriate bioinformatics tools, KEGG was used to
322 obtain the names and EC codes for the enzymes corresponding to the metabolic reactions represented in the entire metabolic network (Letunic *et al.*, 2008; Yamada *et al.*, 2011). Ad-
324 hoc computer programs written in R and Perl (http://www.perl.org/) were used for obtaining that information from local copies of KEGG. The initial EC code assignment was done

keeping the information from the originating organism. To explore the variation on the full metabolic network that represents the metabolic activities of the microbiome, we have pulled them so that the same EC code originating from different organisms in each of the considered sample was considered to be linked to the same metabolic function. This information was further integrated with a generic metabolic map, constructed using information from KEGG (Letunic *et al.*, 2008; Yamada *et al.*, 2011) that provides an overview of the complete metabolism in biological systems.

Cell counts

Cell counts were performed over DAPI stained samples immobilized onto 0.22 um-Pore-Size filters by manually cell counting and marking using the Cell Counter plugin incorporated into ImageJ v1.46 (Rasband *et al.*, 1997-2012, http://imagej.nih.gov/ij/) or through automated counting of single color images, also with ImageJ v1.46 (depending on the sample's characteristics). To achieve statistical signification, around 50 fields were checked per sample. Cell numbers were obtained by referring the counts to the screened area and the volume of sample used.

Biochemical experimental setting: determination of glycosidase activity

Fresh samples (25 ml) were collected from B1 streamer at different depths (1, 3, 5, 7, 9 and 15 cm), and stored at +4°C until further processing (within the following 2 h). Protein extracts were isolated as described previously (Alcaide *et al.*, 2012; Hernández *et al.*, 2013). Briefly, the samples (25 ml) were centrifuged at 13,000 *g* for 5 min to pellet the bacterial cells (together with biofilm), which were lysed in BugBuster® Protein Extraction Reagent (Novagen, Darmstadt, Germany) for 30 min at 4°C, with further disruption by sonication (using a pin Sonicator® 3000; Misonix) for 2.5 min (10 watts) on ice (5 cycles x 0.5 min). The

extracts were then centrifuged for 10 min at 12,000 *g* to separate biofilm material, cell debris and intact cells and the supernatants carefully aspirated (to avoid disturbing the pellet) and transferred to a new tube. Proteins were stored at -80°C at a ~1.0 mg/ml concentration until use. Glycosidase activity was quantified in triplicates in 96-well plates using a BioTek Synergy HT spectrophotometer in a colorimetric assay with a set of 14 structurally diverse sugars (as *p*-nitrophenol [*p*NP] derivatives, all from Sigma Chemical Co., St. Louis, MO, USA) and under previously described conditions (Alcaide *et al*., 2012; Hernández *et al*., 2013), with minor modifications. Briefly, reactions contained 4.0 µg total protein and 1 mg/ml sugar substrate (from a 10-mg/ml stock solution) in a 20-mM 4-(2-hydroxyethylpiperazine-1-ethanesulfonic acid (HEPES) buffer (pH 7.0, *T*=30°C in a final volume of 150 µl). Reactions were followed every 2 min by measuring the release of *p*NP at 410 nm for a total of 130-min. In all cases, one unit (U) of enzyme activity was defined as the amount of protein producing 1 µmol of reducing sugars for 1 min under the assay conditions. Note: quantification of glycosidase activity at different depths was not performed in B2 micro-environment as this is a non-stratified mat.

Supplementary Results and Discussion

Bacterial stratification: clusters description

The following clusters were observed in Fig. 2. Cluster 1 was composed of candidate division WS6-related OPUs (only in B2). Cluster 2 was composed of candidate division TM7-related bacteria. Cluster 3 was composed of *Rhizobium*-related *α-Proteobacteria* (only in B1A). Cluster 4 was composed of *Acidithiobacillus*-related *γ-Proteobacteria* (in B1A and B2). Cluster 5 was related to the *Acidithiobacillales* "RCP1-48" cluster (in B1B as a major constituent and B2), likely representing a new species of the genus *Acidithiobacillus* (*γ-Proteobacteria*). Cluster 6 contained *Xanthomonadaceae*-related *γ-Proteobacteria* (in all

samples). Cluster 7 contained "*Ferrovum myxofaciens"* TRA3-20-related *β-Proteobacteria* (in B1A and B2). Cluster 8 contained *Acidobacterium*-related *Acidobacteria* (in B1B and B2). Cluster 9 contained *Leptospirillum "ferrodiazotrophum*"-related *Nitrospirales* (in B1A). Cluster 10 contained *Leptospirillum* "group I"-related *Nitrospirales* (the group of *Leptospirillum* members affiliated with the characterised species of *Leptospirillum ferrooxidans*) (in B1B and B2). Cluster 11 contained *Clostridiales* (Family XVII Incertae Sedis)-related *Firmicutes* (in B2). Cluster 12 contained *Acidimicrobiaceae*-related *Actinobacteria* (in B1B and B2). Lastly, cluster 13 contained *Actinobacteria*-like bacteria relevant to clone cluster "TM214" (in B1B and B2).

Cell motility and signal transduction proteomic shifts responding to microbial stratification

It appears that the significant enrichment in 'Cell motility' and to a lower extent in 'Signal transduction' related genes, such as those encoding two-component, chemotaxis, flagellar and efflux systems (PhoR, FliCHI, CheABRVYW, Mcp, PleD, OmpR, NarL, UhpB, NtrC, Aer, MotAB and CusB) in the uppermost B1A layer is implicated in the observed differences. Indeed, 8 cell division (FtsZ, FtsE, FtsH and FtsX-like) and 5 chemotaxis (CheY, CheA, Pas, Pac, CheV-like) proteins were identified as being expressed in B1A community (expression level: from 428 to 15 mmol/mol total protein) through quantitative shotgun metaproteome measurements, whereas only 4 (FtsE, FtsK, FtsY and a methyl-accepting chemotaxis sensory transducer; from 21 to 106 mmol/mol total protein) and 3 (FtsL, FtsX, FtsZ; from 67 to 741 mmol/mol total protein) were expressed in B1B and B2, respectively (**Table S6**).

Community formation dynamics also supports a high diversity of biofilm-forming bacteria

FISH (**Fig. S5**) and taxonomic profiling through SSU rRNA tag pyro-sequencing (**Table S4** and **Fig. S6**) analyses were deployed to evaluate the dynamics of microbial population in the

B1 filamentous growths. To that end, we compared bacterial (as no archaeal were detected) populations that have the ability to colonize glass slides submerged in the proximity of the thick B1 streamer at different times (1 hour, 6 hours, and 1, 2, 4, 10, 23 and 58 days). There were several notable findings. First, colonization increased with time (**Fig. S5**). Second, the most abundant first colonizers include *Proteobacteria* (57-96% total reads; **Fig. S6**), namely *β-Proteobacteria* (56-86%), *Acidithiobacillus*-related *γ-Proteobacteria* (3-28%), *Xanthomonadaceae*-related *γ-Proteobacteria* (up to 4%), *Acidithiobacillales* "RCP1-48" related *γ-Proteobacteria* (< 1%), *Rhodospirillales*- (up to 2%) and *Sphingomonadales*- (0.02-1.9%) related *α-Proteobacteria*. *Nitrospirae* (1-22%; maximal at day 10), *Actinobacteria* (1.5 to 10%; maximal at days 10 and 23), *Firmicutes* (0.1-3%; maximal at day 10) and unclassified *Bacteria* (maximal at day 10) were also found among the first colonizers (**Fig. S6**). *Proteobacteria* (57-96%) predominate in all samples (including in mature biofilms), and seem to be the main first colonizers (6 h) (**Fig. S5**). Fourth, the proportion of *Acidobacteria* (0.1-0.4%) slightly increases in late developmental stages. Fifth, phyla that appeared uniquely at a certain sampling times and are not likely to contribute to the community settlement include candidate division AD3 (Zhou *et al*. 2003), *Deinococcus-Thermus* (Weisburg *et al*., 1989) and *Cyanobacteria* representatives, which appear at 6 h (~1%) and 10 days (< 1%); additionally, TM7 candidate division representatives were only perceived at 6 hours, 1 day and 4 days (< 1%) and do not appear until the community is completely structured. Sixth, at 10 days, a drastic community compositional change occurs; thus, *β-Proteobacteria* (56-86%, most remained unclassified) appear to dominate in all samples, except at day 10 (18%), where a higher number of SSU rRNA tags with taxonomic affiliation to *γ-Proteobacteria*, *Nitrospirae* and *Actinobacteria* have been recovered in comparison with previous sampling times (**Fig. S6**). Despite this apparent de-stabilization in the community, the *β-Proteobacteria* dominance becomes manifest again at 58 days, suggesting that the community needs a longer

426 period to get established. Although it is difficult to ascertain whether or not this behavior responds to environmental changes, the fact that geochemical conditions do not vary
428 substantially in short periods of time (**Fig. S2**) suggests that the observed changes within the community structure may be intrinsic to the proper community evolution dynamics.

430 Whatever the evolution of the population, there are clear differences with other AMD systems where the diversity dynamics has been also evaluated: (1) *Tinto River* studied
432 biofilms thrive at an open-air AMD, with significantly different geochemical conditions (Aguilera *at al*. 2007) and with eukaryotes accounting for an important proportion of those
434 communities; (2) the Richmond mine biofilms are *Leptospirillum ferriphilum*-related ("group II")-dominated (Wilmes *et al*. 2009), a group which has not been detected in the *Los Rueldos*
436 formations; and (3) although, most bacteria that rapidly colonized the North Wales channel (*Acidithiobacillus ferrivorans* and "*Ferrovum myxofaciens*"; Johnson, 2012) have been
438 detected through phylogenetic analysis in B1A and B1B, the *Los Rueldos* communities displayed more diversity at both early and mature developmental stages (**Tables S3** and **S4**).

440 In order to evaluate if possible micro-particles of pyrite - the predominant mineral in *Los Rueldos* (abandoned in the 1970s) - could act as attachment surface for the initial community
442 settlement, an evaluation of the pyrite content in the three macroscopic structures was performed through X-ray photoelectron spectroscopy analysis (XPS). Results revealed that
444 for Fe $2p_{3/2}$ structures there were three components, namely FeO at 709.8 eV, FeO(OH) at 712.3 eV and satellite of Fe(II) at 715.6 eV. The absence of a component at 707.3 eV
446 indicates that pyrite is not present or, at least, it is below the detection limit (**Fig. S7**). For the S 2p, the spectrum has been adjusted with doubletons S $2p_{3/2}$- S $2p_{1/2}$, with a fixed intensities
448 relation of 2:1 and a separation of 1.2 eV. The component S $2p_{3/2}$ between 168.5 and 169.0 eV is indicative for the presence of sulphates; the corresponding component for the pyrite, which
450 would have been located at 162.7 eV, is also missing (**Fig. S7**). Taken together, the apparent

absence of pyrite, which fuels the overall process of AMD production in the evaluated samples, leads to the hypothesis that the initial settlement of the communities could occurr from the bank - where the mineral is present - to the stream center. This has also been described by Wilmes *et al*. (2009) in stream biofilms thriving along the Richmond mine´s AMD, where early developmental stages were located in the stream central area, while later developmental stages appeared by the AMD bank, revealing that those regions originated earlier.

Sugar metabolism: genomic, proteomic and turnover shifts responding to microbial stratification

We investigated the presence of additional signatures for sugar metabolism in the community genomes. Out of a total of 57,815 sequences, 262 (B1A: 116, B1B: 65, and B2: 81) glycoside-hydrolases (GH) and carbohydrate-binding modules (CBM) from 32 distinct families and 180 (B1A: 90; B1B: 50; B2: 40) glycoside-transferases (GT) from 17 distinct families, were identified. Abundance-profiling analysis showed overall similar distributions among families (not shown), with an over-representation of GH/CBM classified into the families 94, 13, 15 and 3, in this order, all accounting > 40% of the total GH/CBM. Among GT, those from families 2, 4 and 51 account > 62% of the total, in all samples. This may explain that both B1 and B2 macroscopic growths were composed by similar simple sugars compositions, namely, rhamnose, mannose, glucose and galactose at a molar ratio of 2:1:2:0.36 (for B1) and 2:1:3:0.35 (for B2), as determined by GC/MS. This is of special significance given that both micro-environments contain different microbial populations (**Figs. 2** and **4**), which is indicative of functional redundancy at the level of carbohydrate metabolism.

Among the 442 sugar modifying enzymes tentatively identified through community genome analysis, 36 (B1A: 26, B1B: 15 and B2: 12) were unambiguously identified as being

476 expressed by shotgun proteomics. Only 7 (or 19% total) were shared among the three communities. The expression level was 10 times higher in the B1A oxic stratum as compared
478 to B1B and B2 suboxic mats. Proteins being expressed in B1A support a more diverse enzyme spectrum that included 8 GH proteins from the families 5, 13, 18, 57 and 65, and one
480 rhamnosidase. By contrast, 2 oligosaccharide phosphorylases and 2 GH from families 3 and 15 (in B1B) and 3 GH from families 1 and 20 and 1 exopolysaccharide biosynthesis protein
482 (GT from family 39) (in B2) were expressed in the suboxic mats. Common capacities included those supported by GT from families 1 and 2 (most abundant in all communities) as
484 well as by GH proteins from family 42, α-arabinofuranosidases and polysaccharide biosynthesis proteins/deacetylases of the carboxyl-esterase family 4. So, although the
486 sequence analysis suggests partial redundancy, the proteomic analysis points that there are fundamental differences in the expression of GH and GT, which may have implications in the
488 energy uptake and the growth preferences of B1A populations. In relation to that, the fact that clear shifts in carbohydrate turnovers exist in B1 streamer at 5 cm depth (**Fig. S11**), where a
490 drop in the dissolved oxygen concentration occurs (**Fig. 1**), agrees with the important specific role of this group of enzymes in the biolfilm production and microbial stratification.

492

Supplementary References

494 Aguilera A, Zettler E, Gómez F, Amaral-Zettler L, Rodríguez N, Amils R. (2007). Distribution and seasonal variability in the benthic eukaryotic community of Río Tinto
496 (SW, Spain), an acidic, high metal extreme environment. Syst Appl Microbiol 30:531-546.

498 Alcaide M, Messina E, Richter M, Bargiela R, Peplies J, Huws SA, *et al*. (2012). Gene sets for utilization of primary and secondary nutrition supplies in the distal gut of endangered
500 Iberian lynx. PLoS One 7:e51521.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402

Baker BJ, Tyson GW, Webb RI, Flanagan J, Hugenholtz P, Allen EE, Banfield JF. (2006). Lineages of acidophilic archaea revealed by community genomic analysis. Science 314:1933-1935.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, *et al*. (2004). The Pfam protein families database. Nucleic Acids Res 32:D138-141.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. (2004). Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783-795.

Bradford MM. (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248-254.

Brown JF, Jones DS, Mills DB, Macalady JL, Burgos WD. (2011). Application of a depositional facies model to an acid mine drainage site. Appl Environ Microbiol 77:545-554.

Bruneel O, Duran R, Casiot C, Elbaz-Poulichet F, Personné JC. (2006). Diversity of microorganisms in Fe-As-rich acid mine drainage waters of Carnoulès, France. Appl Environ Microbiol 72:551-556.

Bruneel O, Pascault N, Egal M, Bancon-Montigny C, Goñi-Urriza MS, Elbaz-Poulichet F, Personné JC, Duran R. (2008). Archaeal diversity in a Fe-As rich acid mine drainage at Carnoulès (France). Extremophiles 12:563-571.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, *et al*. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335-336.

526 Chen VB, Davis IW, Richardson DC. (2009). KING (Kinemage, Next Generation): a versatile interactive molecular and scientific visualization program. Protein Sci 18:2403-2409.

528 Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, *et al*. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids 530 Res 37:D141-145.

DeLong EF. (1992). Archaea in coastal marine environments. Proc Natl Acad Sci USA 532 89:5685-5689.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, *et al*. (2006). 534 Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069-5072.

536 Druschel GK, Baker BJ, Gihring TM, Banfield JF. (2004). Acid mine drainage biogeochemistry at Iron Mountain, California. Geochem T 5:13-32.

538 Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-2461.

540 Fierer N, Liu Z, Rodríguez-Hernández M, Knight R, Henn M, Hernandez MT. (2008). Short-term temporal variability in airborne bacterial and fungal populations. Appl Environ 542 Microbiol 74:200-207.

González-Toril E, Llobet-Brossa E, Casamayor EO, Amann R, Amils R. (2003). Microbial 544 ecology of an extreme acidic environment, the Tinto River. Appl Environ Microbiol 69:4853-4865.

546 Hao C, Wang L, Gao Y, Zhang L, Dong H. (2010). Microbial diversity in acid mine drainage of Xiang Mountain sulfide mine, Anhui Province, China. Extremophiles 14:465-474.

548 Hongoh Y, Yuzawa H, Ohkuma M, Kudo T. (2003). Evaluation of primers and PCR conditions for the analysis of 16S rRNA genes from a natural environment. FEMS 550 Microbiol Lett 221:299-304.

Huber T, Faulkner G, Hugenholtz P. (2004). Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20:2317-2319.

Huse SM, Huber JA, Morrison HG, Sogin ML, Weich DM. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 8:R143.

Johnson DB, Hallberg KB. (2003). The microbiology of acidic mine waters. Res Microbiol 154:466-473.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, *et al*. (2008). KEGG for linking genomes to life and the environment. Nucleic Acids Res 36:D480-D484.

Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567-580.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ Microbiol 12:118-123.

Lane DJ. (1991). 16S/23S rRNA sequencing. *In*: Nucleic acid techniques in bacterial systematics. Stackebrandt E, and Goodfellow M, eds., John Wiley and Sons, New York, pp. 115-175.

Letunic I, Yamada T, Kanehisa M, Bork P. (2009). iPath: interactive exploration of biochemical pathways and networks. Trends Biochem Sci 33:101-103.

Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Jacob B, *et al*. (2012). IMG/M: the integrated metagenome data management and comparative analysis system. Nucleic Acids Res 40:D123-129.

Massana R, Murray AE, Preston CM, DeLong EF. (1997). Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. Appl Environ Microbiol 63:50-56.

576 Masuda T, Tomita M, Ishihama Y. (2008). Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. J Proteome Res 7:731-740.

578 Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, *et al*. (2003). GenDB - an open source genome annotation system for prokaryote genomes. Nucleic Acids Res 580 31:2187-2195.

Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns P, *et al*. (2005). InterPro, 582 progress and status in 2005. Nucleic Acids Res 33:D201-205.

Needleman SB, Wunsch CD. (1970). A general method applicable to the search for 584 similarities in the amino acid sequence of two proteins. J Mol Biol 48:443-453.

Noguchi H, Park J, Takagi T. (2006). MetaGene: prokaryotic gene finding from 586 environmental genome shotgun sequences. Nucleic Acids Res 34:5623-5630.

Oksanen J, Guillaume Blanchet F, Kindt R, Legendre P, Minchin PR, O'Hara RB, *et al*. 588 (2010). Vegan: Community Ecology Package. R package version 2.0-5.

Ovreås L, Forney L, Daae FL, Torsvik V. (1997). Distribution of bacterioplankton in 590 meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. Appl Environ Microbiol 592 63:3367-3373.

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2--approximately maximum-likelihood trees 594 for large alignments. PLoS One 5:e9490.

Pruesse E, Peplies J, Glockner FO. (2012). SINA: accurate high-throughput multiple 596 sequence alignment of ribosomal RNA genes. Bioinformatics 28:1823-1829.

Quast C. (2006). MicHanThi - design and implementation of a system for the prediction of 598 gene functions in genome annotation projects. Master Thesis; available on request.

Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, 600 http://imagej.nih.gov/ij/, 1997-2012.

Raskin L, Stromley JM, Rittmann BE, Stahl DA. (1994). Group-specific 16S rRNA
602 hybridization probes to describe natural communities of methanogens. Appl Environ Microbiol 60:1232-1240.

604 Richter M, Lombardot T, Kostadinov I, Kottmann R, Duhaime MB, Peplies J, Glöckner FO. (2008). JCoast - a biologist-centric software tool for data mining and comparison of
606 prokaryotic metagenomes. BMC Bioinformatics 9:177.

Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ. (2006). Absolute quantification
608 of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics 5:144-156.

610 Tan GL, Shu WS, Hallberg KB, Li F, Lan CY, Huang LN. (2007). Cultivation-dependent and cultivation-independent characterization of the microbial community in acid mine
612 drainage associated with acidic Pb/Zn mine tailings at Lechang, Guangdong, China. FEMS Microbiol Ecol 59:118-126.

614 Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, *et al*. (2001). The COG database: new developments in phylogenetic classification of proteins
616 from complete genomes. Nucleic Acids Res 29:22-28.

Viollier E, Inglett PW, Hunter K, Roychoudhury AN, Van Cappellen P. (2000). The ferrozine
618 method revisited: Fe(II)/Fe(III) determination in natural waters. Appl Geochem 15:786-796.

620 Weisburg WG, Giovannoni SJ, Woese CR. (1989). The *Deinococcus-Thermus* phylum and the effect of rRNA composition on phylogenetic tree construction. Syst Appl Microbiol
622 11:128-34.

Wickham H. ggplot2: elegant graphics for data analysis. Springer New York, 2009.

624 Xiao S, Xie X, Liu J, He Z, Hu Y. (2008). Compositions and structures of archaeal communities in acid mineral bioleaching systems of Dongxiang Copper Mine and
626 Yinshan Lead-Zinc Mine, China. Curr Microbiol 57:239-244.

Xie X, Xiao S, He Z, Liu J, Qiu G. (2007). Microbial populations in acid mineral bioleaching
628 systems of Tong Shankou Copper Mine, China. J Appl Microbiol 103:1227-123

Yamada T, Letunic I, Okuda S, Kanehisa M, Bork P. (2011). iPath2.0: interactive pathway
630 explorer. Nucleic Acids Res 39:W412-W415.

632 **Supplementary Tables**

Table S1 Geochemical parameters of the *Los Rueldos* AMD site in comparison with other
634 AMDs. Values are represented in mg/L. References: (i) *Storwartz* (copper), *Ynysarwed* (coal), Wheal Jane (tin), *King's* stream (copper) and *Parys* (copper) mines (Johnson and Hallberg
636 2003; Hallberg *et al*., 2006); (ii) *Xiang* (sulphide) mine (Hao *et al*., 2010); (iii) Carnoulès (lead-zinc) mine (Bruneel *et al*. 2006, 2008; Bertin *et al*., 2011). (iv) *Tong Shankou* (copper)
638 mine (Xie *et al*., 2007); (v) *Tinto River* (copper; also manganese and silver) (Johnson and Hallberg, 2003; González-Toril *et al*., 2003; López-Archilla *et al*., 2004; García-Moyano *et*
640 *al*., 2007; Amils *et al*., 2011; González-Toril *et al.*, 2011); (vi) *Yinshan* (lead-zinc) mine (He *et al*., 2008; Xiao *et al*., 2008; Yin *et al*., 2008); (vii) *Dongxiang* (copper) mine (He *et al*.,
642 2008; Xiao *et al*., 2008); (viii) *Lechang* (lead-zinc) mine (Tan *et al*., 2007); (ix) *Iron Mountain* (iron, silver, gold, copper, zinc, pyrite) (Bond *et al*., 2000; Baker and Banfield,
644 2003; Druschel *et al*., 2004); (x) *Harz-Mountains* (pyrite) mine (Ziegler *et al*., 2009, 2013; Johnson, 2012); and (xi) *Lower Red Eyes* (coal) mine (Brown *et al*., 2011).

646

Table S2 Blueprints of the microbial diversity and composition of the B1A, B1B and B2
648 communities. Tentative phylogenetic affiliations are specifically indicated.

Table S3 Statistical microbial diversity indices.

Table S4 Statistical bacterial diversity indexes during the follow-up colonization assay.

Table S5 General features of the B1A, B1B and B2 metagenomes.

Table S6 Proteins of sampled populations identified and quantified by metaproteomic approaches. Protein annotation is specifically shown. For raw data, please contact authors directly. The Table is provided as separate file (Excel file) because it contains massive information.

Supplementary Figures

Fig. S1 Geographic situation of the *Los Rueldos* emplacement. Locations of the Hg spoil heap and cave are shown.

Fig. S2 Statistical analysis performed over a set of geochemical values collected from June 2010 to July 2011 in the two discrete micro-environments identified within the *Los Rueldos* ecosystem. **(A)** The evolution of pH, Eh, temperature and conductivity during the sampling period. **(B)** Comparison of mean values considering the season as a factor (season 1, January, February, March; season 2, the rest of the year). **(C)** One-way ANOVA results displaying significant differences in the values for the variables measured in seasons 1 and 2. **(D)** Correlation analysis for each of the variables. Eh values are given in mV, temperature in °C and conductivity in mS/cm.

Fig. S3. Overview of the relative abundance of phylogenetic microbial groups recovered from the *Los Rueldos* macroscopic growths based on both 16S rRNA clone libraries and SSU rRNA hypervariable tag pyro-sequencing analyses. **(A)** Phylum, class and order-level taxonomic affiliations for B1A, B1B and B2 bacterial sequences. Columns are arranged by pairs according to the methodology used for obtaining the taxonomic profile: FL, full-length 16S rRNA gene analysis; V1-V3, hypervariable tag pyro-sequencing analysis. Grey boxes next to the phyla-level classifications contain rarefaction plots obtained by the two taxonomic profiling methodologies assayed. **(B)** Phylum, class and order-level taxonomic affiliations for B1A, B1B and B2 archaeal sequences, including ultra-microarchaea. Column charts arranged as specified in (A). Rarefaction plots corresponding to each methodology are presented on the right margin of the figure: x axis, number of sequences per sample; y axis, number of species. Colour coding as follows: B1A, red; B1B, blue; B2, green. Asterisk (*), refer to main text for details about the E2 group of uncultured euryarchaea.

Fig. S4 Maximum likelihood phylogenies constructed from the obtained full-length 16S rRNA gene "ARMAN"-related sequences. The main phylogenetic clusters are highlighted in grey. Bootstrap values are indicated the corresponding nodes with circles (>85%, closed circles; >50%, open circles), except for the "ARMAN"-related phylogeny, where all bootstrap values are >85%. Scale bars represent changes per site or the (x100) % difference in nucleotide sequences. The contribution of each OPU to the library is represented in the accompanying graphs. Representative 16S rRNA sequences are coloured in accordance with the sample: B1A, red; B1B, blue; B2, green.

Fig. S5 Fluorescence microscopy images the main bacterial morphologies present during the colonization assay. DAPI stained cells (left) and bacterial probe (right) hybridizations at 6 h **(A)**, 2 days **(B)**, 10 days **(C),** 23 days, **(D)** and 58 days **(E)**.

Fig. S6 Overview of the relative abundance of phylogenetic microbial groups implicated in the settlement of the initial B1 community. **(A)** Phylum, class and order-level taxonomic affiliations for the slides recovered bacterial sequences through SSU rRNA hypervariable tag pyro-sequencing analysis in comparison to the corresponding taxonomic profiles obtained for the mature B1 (A and B) and B2 communities. **(B)** Photography of the region where slides were submerged, where it becomes visible the discontinuity of B1 with the slight augmentation of the depth (inset); below, schematic representation of the biofilm-collector device used. **(C)** Temporal evolution in the proportion of the phylum (1), class (2) and order-level (3) taxonomic groups detected at all sampling times. Corresponding colored lines´ assignations are represented next to the taxa indicated in the (A) legends. **(D)** Rarefaction plot corresponding to the analyzed samples.

Fig. S7 X-ray photoelectron spectroscopy spectra for the Fe $2p_{3/2}$ and S 2p components over dried and homogenized B1A, B1B and B2 structures. Asterisk (*), arbitrary units.

Fig. S8 Expression patterns and distributions of taxonomic categories of proteins that were identified in the B1, B1B and B2 metaproteomes. Expression pattern refers to the absolute protein abundance (mol protein/mol total protein) in each of the samples.

Fig. S9 Community behavior characterized by mapping to metabolic pathways. Proteome-scale metabolic reconstructions were based on the expression values of a set of proteins with

Enzyme Commission (EC) numbers that were found among the 1,589 proteins that were unambiguously quantified in the metaproteomes. Ad-hoc computer programs written in C and Perl were used to obtain information from local copies of KEGG and to generate the graphical representations. Panels (**A**), (**B**) and (**C**) represent B1A, B1B and B2 community profiles, respectively. This Figure (as three independent panels) is provided as separate file (scalable PDF version) to ensure high quality resolution.

Fig. S10 Heat map and hierarchical cluster of the samples based on the relative contribution of each metabolic pathway. The contribution to each pathway was computed based on the average relative weight of the reactions linked to a given pathway present in each sample. KEGG identifiers used for the pathways are specifically shown. Codes as follows: ec00010 Glycolysis / Gluconeogenesis; ec00020 Citrate cycle (TCA cycle); ec00030 Pentose phosphate pathway; ec00040 Pentose and glucuronate interconversions; ec00051 Fructose and mannose metabolism; ec00052 Galactose metabolism; ec00053 Ascorbate and aldarate metabolism; ec00061 Fatty acid biosynthesis; ec00062 Fatty acid elongation; ec00071 Fatty acid metabolism; ec00072 Synthesis and degradation of ketone bodies; ec00120 Primary bile acid biosynthesis; ec00130 Ubiquinone and other terpenoid-quinone biosynthesis; ec00140 Steroid hormone biosynthesis; ec00190 Oxidative phosphorylation; ec00195 Photosynthesis; ec00230 Purine metabolism; ec00240 Pyrimidine metabolism; ec00250 Alanine, aspartate and glutamate metabolism; ec00260 Glycine, serine and threonine metabolism; ec00270 Cysteine and methionine metabolism; ec00280 Valine, leucine and isoleucine degradation ; ec00281 Geraniol degradation; ec00290 Valine, leucine and isoleucine biosynthesis; ec00300 Lysine biosynthesis; ec00310 Lysine degradation; ec00311 Penicillin and cephalosporin biosynthesis; ec00312 beta-Lactam resistance; ec00330 Arginine and proline metabolism; ec00340 Histidine metabolism; ec00350 Tyrosine metabolism; ec00360 Phenylalanine

metabolism; ec00361 Chlorocyclohexane and chlorobenzene degradation; ec00362 Benzoate degradation; ec00364 Fluorobenzoate degradation; ec00380 Tryptophan metabolism; ec00400 Phenylalanine, tyrosine and tryptophan biosynthesis; ec00410 beta-Alanine metabolism; ec00430 Taurine and hypotaurine metabolism; ec00450 Selenocompound metabolism; ec00460 Cyanoamino acid metabolism; ec00471 D-Glutamine and D-glutamate metabolism; ec00473 D-Alanine metabolism; ec00480 Glutathione metabolism; ec00500 Starch and sucrose metabolism; ec00511 Other glycan degradation; ec00520 Amino sugar and nucleotide sugar metabolism; ec00521 Streptomycin biosynthesis; ec00523 Polyketide sugar unit biosynthesis; ec00524 Butirosin and neomycin biosynthesis; ec00531 Glycosaminoglycan degradation; ec00540 Lipopolysaccharide biosynthesis; ec00550 Peptidoglycan biosynthesis; ec00561 Glycerolipid metabolism; ec00564 Glycerophospholipid metabolism; ec00592 alpha-Linolenic acid metabolism; ec00600 Sphingolipid metabolism; ec00604 Glycosphingolipid biosynthesis - ganglio series; ec00620 Pyruvate metabolism; ec00622 Xylene degradation; ec00623 Toluene degradation; ec00625 Chloroalkane and chloroalkene degradation; ec00626 Naphthalene degradation; ec00627 Aminobenzoate degradation; ec00630 Glyoxylate and dicarboxylate metabolism; ec00633 Nitrotoluene degradation; ec00640 Propanoate metabolism; ec00643 Styrene degradation; ec00650 Butanoate metabolism; ec00660 C5-Branched dibasic acid metabolism; ec00670 One carbon pool by folate; ec00680 Methane metabolism; ec00710 Carbon fixation in photosynthetic organisms; ec00720 Carbon fixation pathways in prokaryotes; ec00730 Thiamine metabolism; ec00740 Riboflavin metabolism; ec00750 Vitamin B6 metabolism; ec00760 Nicotinate and nicotinamide metabolism; ec00770 Pantothenate and CoA biosynthesis; ec00780 Biotin metabolism; ec00790 Folate biosynthesis; ec00830 Retinol metabolism; ec00860 Porphyrin and chlorophyll metabolism; ec00900 Terpenoid backbone biosynthesis; ec00903 Limonene and pinene degradation; ec00910 Nitrogen metabolism; ec00920 Sulfur metabolism; ec00930

772 Caprolactam degradation; ec00940 Phenylpropanoid biosynthesis; ec00941 Flavonoid biosynthesis; ec00966 Glucosinolate biosynthesis; ec00980 Metabolism of xenobiotics by
774 cytochrome P450; ec00982 Drug metabolism - cytochrome P450; ec00983 Drug metabolism - other enzymes; ec01040 Biosynthesis of unsaturated fatty acids; ec01051 Biosynthesis of
776 ansamycins. The colour scale (defined as colour key in the figure) quantifies the relative percentages from one sample to the other, and goes from red (highly abundant) to black
778 (abundant) to light green (low abundant).

780 **Fig. S11.** Shifts in carbohydrate turnover profile in the B1 streamer. Enzyme activities (units/g total protein) from the samples obtained from the B1 streamer at different depths
782 against 14 different sugar substrates (as *p*NP derivatives) were quantified by measuring the release of *p*NP in triplicates. A single plot, with mean values estimated for each group of
784 samples using three independent measurements each (SD lower than 5%) is shown.

786 **Fig. S12.** Contigs length distribution of B1A (A), B1B (B) and B2 (C) metagenomes.

Table S1 Geochemical parameters of the *Los Rueldos* AMD site in comparison with other AMDs. Values are represented in mg/L. References: (i) *Storwartz* (copper), *Ynysarwed* (coal), Wheal Jane (tin), *King's* stream (copper) and *Parys* (copper) mines (Johnson and Hallberg 2003; Hallberg *et al*., 2006); (ii) *Xiang* (sulphide) mine (Hao *et al*., 2010); (iii) Carnoulès (lead-zinc) mine (Bruneel *et al*. 2006, 2008; Bertin *et al*., 2011). (iv) *Tong Shankou* (copper) mine (Xie *et al*., 2007); (v) *Tinto River* (copper; also manganese and silver) (Johnson and Hallberg, 2003; González-Toril *et al*., 2003; López-Archilla *et al*., 2004; García-Moyano *et al*., 2007; Amils *et al*., 2011; González-Toril *et al*., 2011); (vi) *Yinshan* (lead-zinc) mine (He *et al*., 2008; Xiao *et al*., 2008; Yin *et al*., 2008); (vii) *Dongxiang* (copper) mine (He *et al*., 2008; Xiao *et al*., 2008); (viii) *Lechang* (lead-zinc) mine (Tan *et al*., 2007); (ix) *Iron Mountain* (iron, silver, gold, copper, zinc, pyrite) (Bond *et al*., 2000; Baker and Banfield, 2003; Druschel *et al*., 2004); (x) *Harz-Mountains* (pyrite) mine (Ziegler *et al*., 2009, 2013; Johnson, 2012); and (xi) *Lower Red Eyes* (coal) mine (Brown *et al*., 2011).

	Storwartz	Ynysarwed	Xiang	Wheal	Carnoulès	King´s	Lower	Mynydd	Tong	Cae Coch	Drei Kronen	Tinto river	Yinshan	Dongxiang	Los Rueldos	Lechang	Richmond
	Norway	*UK*	*China*	*UK*	*France*	*Norway*	*USA*	*UK*	*China*	*UK*	*Germany*	*Spain*	*China*	*China*	*Spain*	*China*	*USA*
pH	6.5	6.2	3	3.4	2.7-3.4	2.75	2.57-4.21	2.5	2.5	1.85-2.30	2.3	2.2	1-2.5	2	2	1.9-2.1	0.5-1
AMD T (^{0}C)	-	-	30	-	-	-	9-28	-	25	8-9	-	-	20-28	20.1	10-15	27-28	15-60
Eh (mV)	-	214	257	532	-	-		685	-	611-796	-	450	-		234-264	689	
Fe_{total}	1.6	160	61	100.6	750-2700	172	304-559	650	545-1300	1005-31570	-	2300	235.5-8040	46890	395-545	4841	13000-19000
Fe^{2+}	-	140	45	0.5	-	-	193-561	650	-	220-11800	-	1500	-		96-140	-	13000-19000
Al	0.03	-	1.2	1186	-	22.5	42-45	70	275.9-284.9	19-265	-	-	136.1-629.3	3471.7	129	-	1400-6700
As	-	-	-	-	100-350	-	-	-	-	-	-	-	0.01-29	9.7	3.6-6	0.22	-
Sb	-	-	-	-	-	-	-	-	-	-	-	-	0.01-7.1		0.013	-	-
Hg	-	-	-	-	-	-	-	-	0.01-0.5	-	-	-	3-3.9		$8.30x10^{-05}$	-	-
Pb	-	-	-	-	-	-	-	-	-	-	-	-	0.05-0.4	21.6	$7.6x10^{-03}$	0.7	-
Mn	1.35	-	15	433.1	-	0.78	109-122	10	-	-	-	-	4.3-207		-	10.9	17-120
Cu	0.06	-	-	51.8	-	15.8	-	40	374.4-990	0.4-37	-	109	33.5-264.6	245.2	-	3	120-650
Zn	2.13	-	-	-	-	25.4	12-15	60	-	13-120	-	225	24-517.5	408.4	-	97	700-2600
Sulfate	151	464	-	1883.6	2.0-7500	668	1892-4294	1550	-	3460-84400	≤19212	10000	552.3-6550	27900	3049-3180	11100	20000-180000

Table S2 Blueprints of the microbial diversity and composition of the B1A, B1B and B2 communities. Tentative phylogenetic affiliations are specifically indicated.

Domain	Sample	Clones	Number of clones	Nearest BLAST match (accession number)	Isolation source of nearest match	Similarity (%)	Phylogenetic affiliation SILVA / RDP
Bacteria	**B1A**	**LR_AB_113**	118	Uncultured bacterium clone RT6-ant09-a11-W (JF737909)	Tinto river (SW Spain)	99	*β-Proteobacteria*
		LR_AB_109	25	*Acidithiobacillus ferrivorans* SS3 (CP002985)	Metal mine impacted environment	99	*γ-Proteobacteria*
		LR_AB_273	7	Uncultured bacterium clone H12 (DQ328620)	Extreme AMD sediment (China)	98	*Nitrospira* (group III)
		LR_AB_238	3	Uncultured bacterium clone CG-36 (FN391831)	As-rich sediments from Carnoulès AMD (France)	99	*γ-Proteobacteria*
		LR_AB_199	3	Uncultured bacterium clone RT11-ant04-c06-S (JF737864)	Tinto river (SW Spain)	99	*Actinobacteria*
		LR_AB_13	2	Uncultured bacterium clone EDW07B001_16 (HM066280)	Texas state well (USA)	99	*α-Proteobacteria*
	B1B	**LR_BB_111**	62	Uncultured bacterium clone RT8-ant03-c06-W (JF807635)	Tinto river (SW Spain)	99	*γ-Proteobacteria*
		LR_BB_102	16	Uncultured bacterium clone BacI_clone9 (FN870221)	Acidic coal mine lake (Germany)	99	*γ-Proteobacteria*
		LR_BB_141	7	Uncultured bacterium clone UMA_h7 (JQ815838)	Tinto river sediment (SW Spain)	97	*Actinobacteria*
		LR_BB_69	2	Uncultured *Acidobacterium* sp. clone JL123_18 (HQ730658)	Anaerobic acidic sediments with heavy metals, Tinto river (SW Spain)	99	*Acidobacteria*
		LR_BB_129	2	Uncultured bacterium clone RT3-ant08-a03-S (JF737901)	Tinto river (SW Spain)	99	*Nitrospira* (group I)
	B2	**LR_2B_268**	23	Uncultured bacterium clone RT6-ant09-a11-W (JF737909)	Tinto river (SW Spain)	98	*β-Proteobacteria*
		LR_2B_1	16	Uncultured bacterium clone RT8-ant03-c06-W (JF807635)	Tinto river (SW Spain)	99	*γ-Proteobacteria*
		LR_2B_74	11	Uncultured bacterium clone B-S-b4 (EU372637)	Tinto river sediment (SW Spain)	99	*Nitrospira* (group I)
		LR_2B_71	8	*Acidithiobacillus ferrooxidans* (JF934687)	Heap leaching plant, Chambishi copper mine (Zambia)	99	*γ-Proteobacteria*
		LR_2B_ 112	8	Uncultured *Acidimicrobiaceae* bacterium SN1_2009_6C (HQ730674)	Anaerobic acidic sediments with heavy metals, Tinto river (SW Spain)	99	*Actinobacteria*
		LR_2B_264	8	Uncultured bacterium clone UMA_h7 (JQ815838)	Tinto river (SW Spain)	96	*Actinobacteria*
		LR_2B_212	6	Uncultured *Acidobacterium* sp. clone JL123_18 (HQ730658)	Anaerobic acidic sediments with heavy metals, Tinto river (SW Spain)	99	*Acidobacteria*
		LR_2B_209	4	Uncultured bacterium clone CG-36 (FN391831)	As-rich sediments from Carnoulès AMD (France)	98	*γ-Proteobacteria*
		LR_2B_263	3	Uncultured bacterium partial clone BacI_clone9 (FN870221)	Acidic coal mine lake (Germany)	98	*γ-Proteobacteria*
		LR_2B_308	3	Uncultured bacterium clone AG_h1 (JQ815634)	Tinto river sediment (SW Spain)	95	unclassified
		LR_2B_137	3	Uncultured bacterium clone RCP1-39 (AF523932)	Forested wetland impacted by reject coal	98	unclassified
		LR_2B_218	3	Uncultured bacterium clone RCP1-88 (AF523933)	Forested wetland impacted by reject coal	98	unclassified
		LR_2B_253	2	Uncultured bacterium clone RT11-ant04-c06-S (JF737864)	Tinto river (SW Spain)	99	*Actinobacteria*
		LR_2B_54	2	Uncultured *Firmicutes* bacterium clone K6-C05 (EF612375)	Lead-zinc mine tailings, Klondyke Mill Site, Arizona (USA)	94	unclassified
Archaea	**B1A**	**LR_AA_9**	41	Uncultured archaeon clone ARCP1-60 (AF523941)	Forested wetland impacted by reject coal	99	unclassified
		LR_AA_47	23	Uncultured archaeon (AB600346)	Acidic spring, Kanagawa (Japan)	99	*Thermoplasmatales*
		LR_AA_20	10	Uncultured archaeon clone antb10 (EF446196)	Floating macroscopic filaments, Tinto river (SW Spain)	98	unclassified/*Thermoplasmatales*

		LR_AA_77	3	Uncultured *Thermoplasmatales* archaeon clone SALE1B1 (FJ228392)	Sediment of acidic mine pit lake (Germany)	92	unclassified
		LR_AA_28	2	Uncultured archaeon clone CEM_Pin_e2 (EU370308)	Tinto river acidic photosynthetic biofilms	99	*Thermoplasmatales*
	B1B	**LR_BA_75**	17	Uncultured archaeon clone CEM_Pin_e2 (EU370308)	Tinto river acidic photosynthetic biofilms (SW Spain)	98	unclassified
		LR_BA_24	10	Uncultured Thermoplasmatales archaeon clone SALE1B1 (FJ228392)	Sediment of acidic mine pit lake (Germany)	95	unclassified
		LR_BA_23	3	Uncultured Thermoplasmatales archaeon clone SALE1A5 (FJ228391)	Sediment of acidic mine pit lake (Germany)	95	unclassified
		LR_BA_84	2	Uncultured archaeon clone HO28S9A63 (AB600346)	Acidic spring, Kanagawa (Japan)	95	unclassified
		LR_BA_55	2	Uncultured archaeon clone anta6 (DQ303248)	Floating macroscopic filaments, Tinto river	99	*Thermoplasmatales*
	B2	**LR_2A_61**	20	Uncultured Thermoplasmatales archaeon clone SALE1B1 (FJ228392)	Sediment of acidic mine pit lake (Germany)	99	*Thermoplasmatales*
		LR_2A_1	12	Uncultured archaeon clone CEM_Pin_e2 (EU370308)	Tinto river (SW Spain)	99	*Thermoplasmatales*
		LR_2A_68	4	Uncultured archaeon clone HO28S9A06 (AB600330)	Acidic spring, Kanagawa (Japan)	99	*Thermoplasmatales*
		LR_2A_91	2	Uncultured archaeon clone HO28S21A41 (AB600370)	Acidic spring, Kanagawa (Japan)	99	*Thermoplasmatales*
		LR_2A_67	2	Uncultured archaeon clone HO28S9A20 (AB600334)	Acidic spring, Kanagawa (Japan)	98	*Thermoplasmatales*
		LR_2A_75	2	Uncultured archaeon clone ARCP1-60 (AF523941)	Forested wetland impacted by reject coal	96	unclassified
Ultra-microarchaea	**B1A short 16S**	**LR_ANC_167**	113	Uncultured archaeon clone ARMAN-HM_4 (JN662468)	Acidic biofilm, Harz-Mountains (Germany)	99	unclassified
		LR_ANC_168	5	Uncultured archaeon clone sp14-9a (FJ379961)	Volcanic sulphur hot springs, Saint Lucia (Lesser Antilles)	89	unclassified
		LR_ANC_79	3	Uncultured archaeon clone ARMAN-HM_29 (JN662470)	Acidic biofilm, Harz-Mountains (Germany)	100	unclassified
		LR_ANC_67	2	Uncultured archaeon clone ARMAN-HM_7 (JN662469)	Acidic biofilm, Harz-Mountains (Germany)	97	unclassified
	B1A long 16S	**LR_ANM_9**	59	Uncultured euryarchaeote ARMAN-1 (AY652726)	Acid mine drainage biofilm, Iron Mountain (California)	93	unclassified
		LR_ANM_29	1	Uncultured euryarchaeote ARMAN-1 (AY652726)	Acid mine drainage biofilm, Iron Mountain (California)	92	unclassified
		LR_ANM_52	1	Uncultured euryarchaeote ARMAN-1 (AY652726)	Acid mine drainage biofilm, Iron Mountain (California)	93	unclassified
		LR_ANM_81	1	Uncultured euryarchaeote ARMAN-1 (AY652726)	Acid mine drainage biofilm, Iron Mountain (California)	90	unclassified
	B1B short 16S	**LR_BNC_99**	82	Uncultured archaeon clone ARMAN-HM_4 (JN662468)	Acidic biofilm, Harz-Mountains (Germany)	100	unclassified
		LR_BNC_90	18	Uncultured archaeon clone sp14-9a (FJ379961)	Volcanic sulphur hot springs, Saint Lucia (Lesser Antilles)	92	unclassified
	B1B long 16S	**LR_BNM_4**	52	Uncultured euryarchaeote ARMAN-1 (AY652726)	Acid mine drainage biofilm, Iron Mountain (California)	93	unclassified
	B2 short 16S	**LR_2NC_27**	92	Uncultured archaeon clone ARMAN-HM_4 (JN662468)	Acidic biofilm, Harz-Mountains (Germany)	100	unclassified
	B2 long 16S	**LR_2NM_75**	30	Uncultured euryarchaeote ARMAN-1 (AY652726)	Acid mine drainage biofilm, Iron Mountain (California)	94	unclassified

Table S3 Statistical microbial diversity indices

	16S rRNA gene clone libraries												SSU rRNA tag pyrosequencing								
	Bacteria			Archaea			ARMAN 400			ARMAN 1000			Bacteria			Archaea			ARMAN 400		
Index	**B1A**	**B1B**	**B2**	**B1A**	**B1B**	**B2**	**B1A**	**B1B**	**B2**	**B1A**	**B1B**	**B2**	**B1A**	**B1B**	**B2**	**B1A**	**B1B**	**B2**	**B1A**	**B1B**	**B2**
Taxa_S	6	5	13	5	5	5	4	2	1	1	1	1									
Individuals	158	91	98	81	34	40	123	100	92	62	52	30									
Dominance_D	0.56	0.50	0.12	0.48	0.35	0.36	0.87	0.71	1	1	1	1	0.27	0.08	0.21	0.33	0.13	0.21	0.53	0.55	0.53
Simpson_1-D	0.41	0.5	0.88	0.52	0.65	0.65	0.15	0.30	0	0	0	0	0.73	0.92	0.79	0.67	0.87	0.79	0.47	0.45	0.47
Shannon_H	0.85	1	2.29	0.95	1.25	1.24	0.37	0.5	0	0	0	0	3.42	5.40	3.73	3.06	3.83	3.24	2.04	1.81	1.92
Evenness_e^H/S	0.39	0.54	0.76	0.51	0.70	0.69	0.36	0.8	1	1	1	1									
Brillouin	0.8	0.90	2.09	0.88	1.08	1.09	0.33	0.45	0	0	0	0	2.30	3.65	2.52	2.07	2.63	2.20	1.40	1.24	1.32
Menhinick	0.48	0.52	1.31	0.56	0.86	0.8	0.36	0.2	0.1	0.13	0.14	0.18	4.8	6.33	4.28	1.94	1.75	1.16	0.88	0.92	0.82
Margalef	0.99	0.89	2.62	0.91	1.13	1.08	0.62	0.22	0	0	0	0	49.86	64.07	38.83	13.97	21.99	7.29	9.52	9.71	9.38
Equitability_J	0.48	0.61	0.9	0.59	0.78	0.77	0.26	0.68	0	0	0	0	0.39	0.59	0.44	0.45	0.5	0.55	0.32	0.28	0.3
Fisher_alpha	1.23	1.14	4.02	1.18	1.62	1.51	0.79	0.35	0.16	0.17	0.18	0.2	101.11	141.40	77.10	22.82	35.09	10.61	13.38	13.76	13.07
Berger-Parker	0.75	0.68	0.23	0.63	0.5	0.5	0.92	0.82	1	1	1	1	0.47	0.20	0.37	0.55	0.22	0.41	0.72	0.74	0.72
Chao-1	6	5	13	5	5	5	4	2	1	1	1	1	842.28	942.22	581.51	152.43	346.08	64	124.43	123.07	118.08

Table S4 Statistical bacterial diversity indexes during the follow-up colonization assay.

	SSU rRNA tag pyrosequencing						
	Bacteria						
Index	**6h**	**1 day**	**2 days**	**4 days**	**10 days**	**23 days**	**58 days**
Dominance_D	0.65	0.54	0.73	0.36	0.075	0.2	0.51
Simpson_1-D	0.35	0.47	0.27	0.64	0.93	0.8	0.5
Shannon_H	1.2	1.7	0.9	2	4	3.27	1.61
Brillouin	0.67	0.94	0.51	1.17	2.24	1.81	0.93
Menhinick	1.11	1.58	0.79	0.74	0.51	0.78	0.84
Margalef	1.63	2.44	1.08	2.17	5.16	4.61	1.63
Equitability_J	0.43	0.51	0.39	0.64	0.93	0.78	0.57
Fisher_alpha	2.46	4.28	1.51	3.61	15.92	12.6	2.46
Berger-Parker	0.8	0.73	0.85	0.53	0.13	0.43	0.7
Chao-1	17	17	5.3	16.5	42	31.2	7.25

Table S5 General features of the B1A, B1B and B2 metagenomes

Parameter	B1A	B1B	B2
Reads	487368	454404	302143
Average length per read	335.34	391.15	374.97
Raw DNA sequences (Mbp)	163.43	177.89	113.29
Size (bp)	14356296	14901926	10850070
Contigs	16397	13236	11600
Average contig size (bps)	875	1125	935
Average GC content (%)	59.72	63.48	62.25
Protein-coding genes (CDS)	22159	19690	15961
Average CDS size (bp)	527	604	555
tRNAs	182	216	167
rRNAs	28	29	30
ORFs with predicted function	23009	20237	16558
Hypothetical	6039	5190	4558
Conserved hypothetical	5804	4861	4153
Assigned to COGs	15700	14410	11141
Assigned to KEGG	11031	9948	7731

Table S6 Proteins of sampled populations identified and quantified by metaproteomic approaches. Protein annotation is specifically shown. For raw data, please contact authors directly.

The Table is provided as separate file (Excel file) because it contains massive information.

Fig. S1 Geographic situation of the *Los Rueldos* emplacement. Locations of the Hg spoil heap and cave are shown.

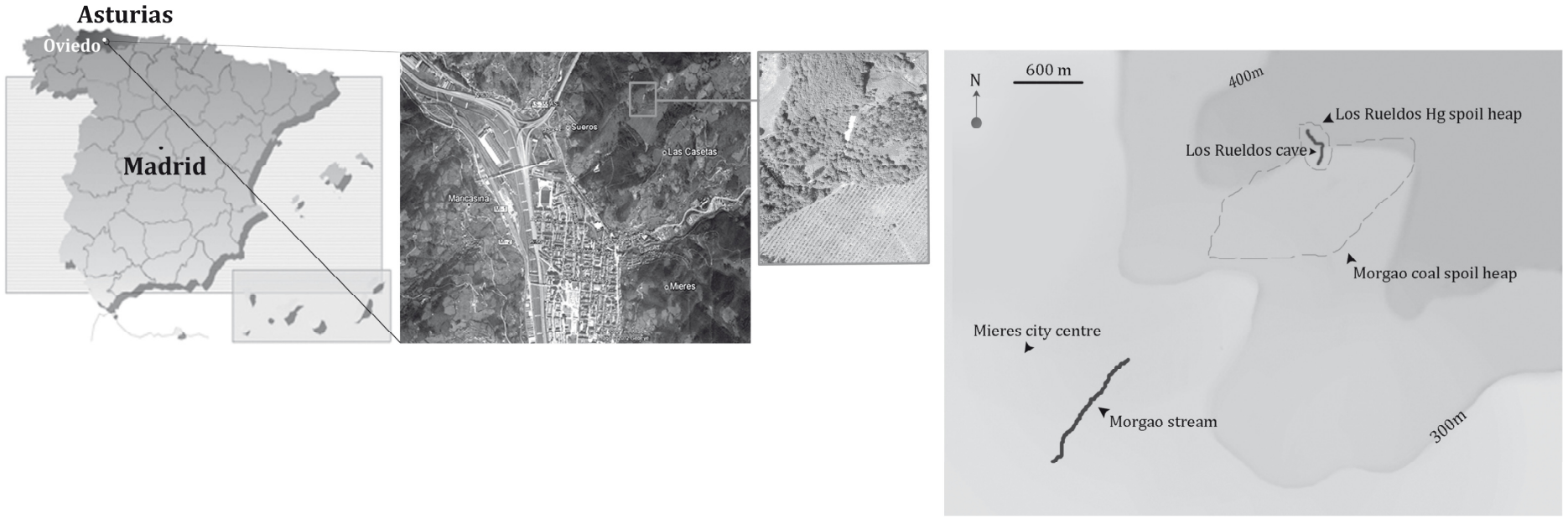


Fig. S2 Statistical analysis performed over a set of geochemical values collected from June 2010 to July 2011 in the two discrete micro-environments identified within the *Los Rueldos* ecosystem. **(A)** The evolution of pH, Eh, temperature and conductivity during the sampling period. **(B)** Comparison of mean values considering the season as a factor (season 1, January, February, March; season 2, the rest of the year). **(C)** One-way ANOVA results displaying significant differences in the values for the variables measured in seasons 1 and 2. **(D)** Correlation analysis for each of the variables. Eh values are given in mV, temperature in °C and conductivity in mS/cm.

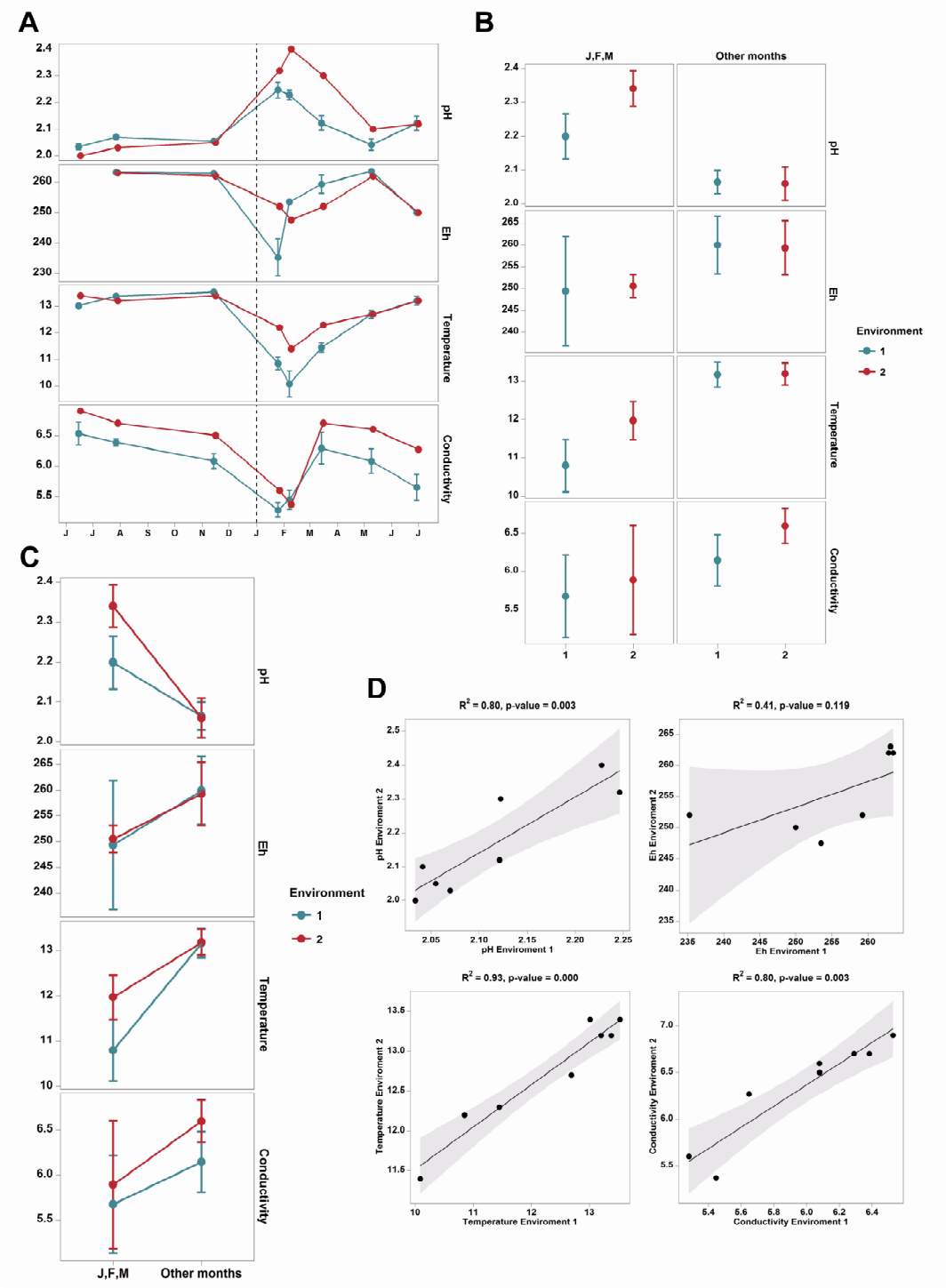


Fig. S3. Overview of the relative abundance of phylogenetic microbial groups recovered from the *Los Rueldos* macroscopic growths based on both 16S rRNA clone libraries and SSU rRNA hypervariable tag pyro-sequencing analyses. **(A)** Phylum, class and order-level taxonomic affiliations for B1A, B1B and B2 bacterial sequences. Columns are arranged by pairs according to the methodology used for obtaining the taxonomic profile: FL, full-length 16S rRNA gene analysis; V1-V3, hypervariable tag pyro-sequencing analysis. Grey boxes next to the phyla-level classifications contain rarefaction plots obtained by the two taxonomic profiling methodologies assayed. **(B)** Phylum, class and order-level taxonomic affiliations for B1A, B1B and B2 archaeal sequences, including ultra-microarchaea. Column charts arranged as specified in (A). Rarefaction plots corresponding to each methodology are presented on the right margin of the figure: x axis, number of sequences per sample; y axis, number of species. Colour coding as follows: B1A, red; B1B, blue; B2, green. Asterisk (*), refer to main text for details about the E2 group of uncultured euryarchaea.

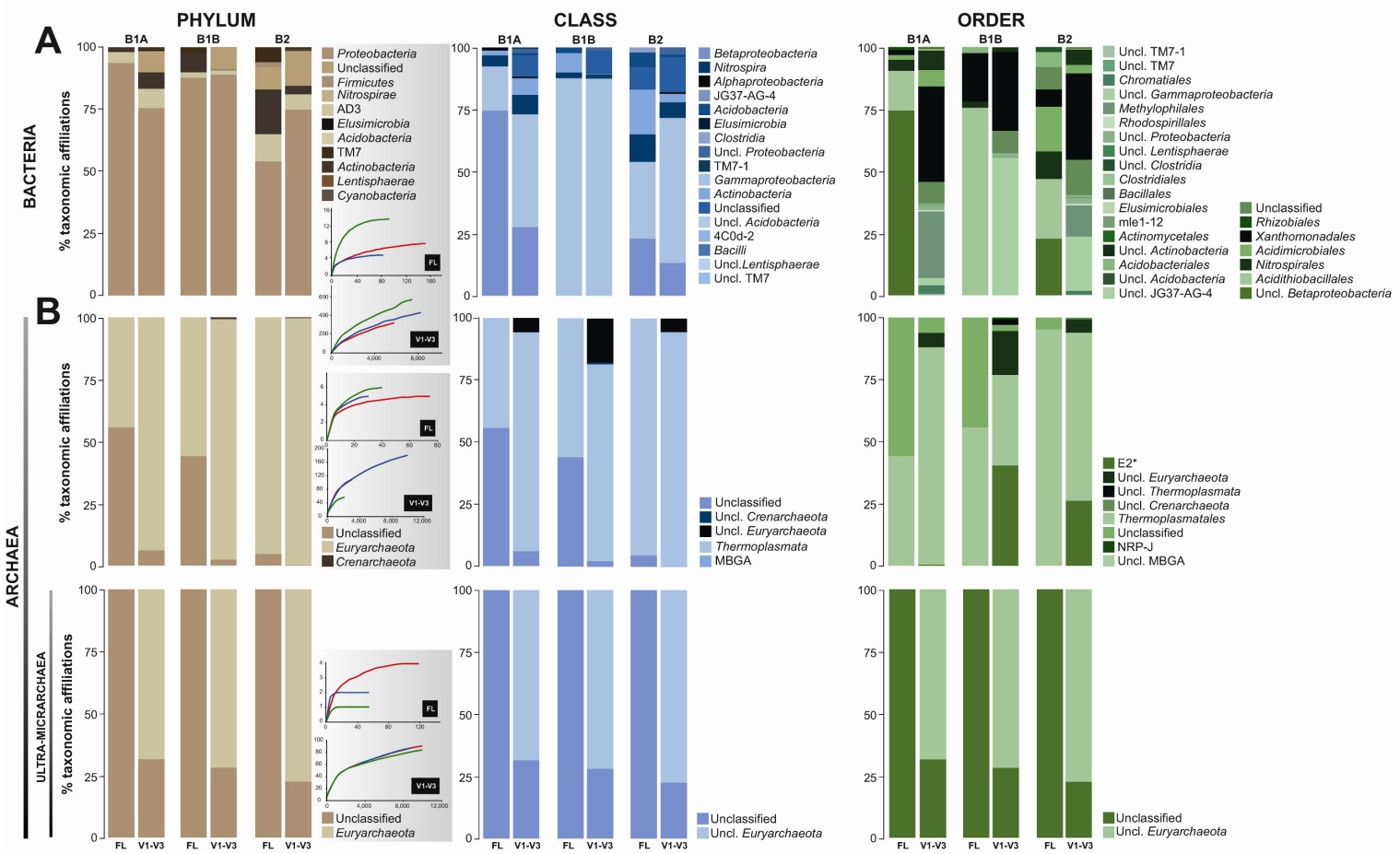


Fig. S4 Maximum likelihood phylogenies constructed from the obtained full-length 16S rRNA gene "ARMAN"-related sequences. The main phylogenetic clusters are highlighted in grey. Bootstrap values are indicated the corresponding nodes with circles (>85%, closed circles; >50%, open circles), except for the "ARMAN"-related phylogeny, where all bootstrap values are >85%. Scale bars represent changes per site or the (x100) % difference in nucleotide sequences. The contribution of each OPU to the library is represented in the accompanying graphs. Representative 16S rRNA sequences are coloured in accordance with the sample: B1A, red; B1B, blue; B2, green.

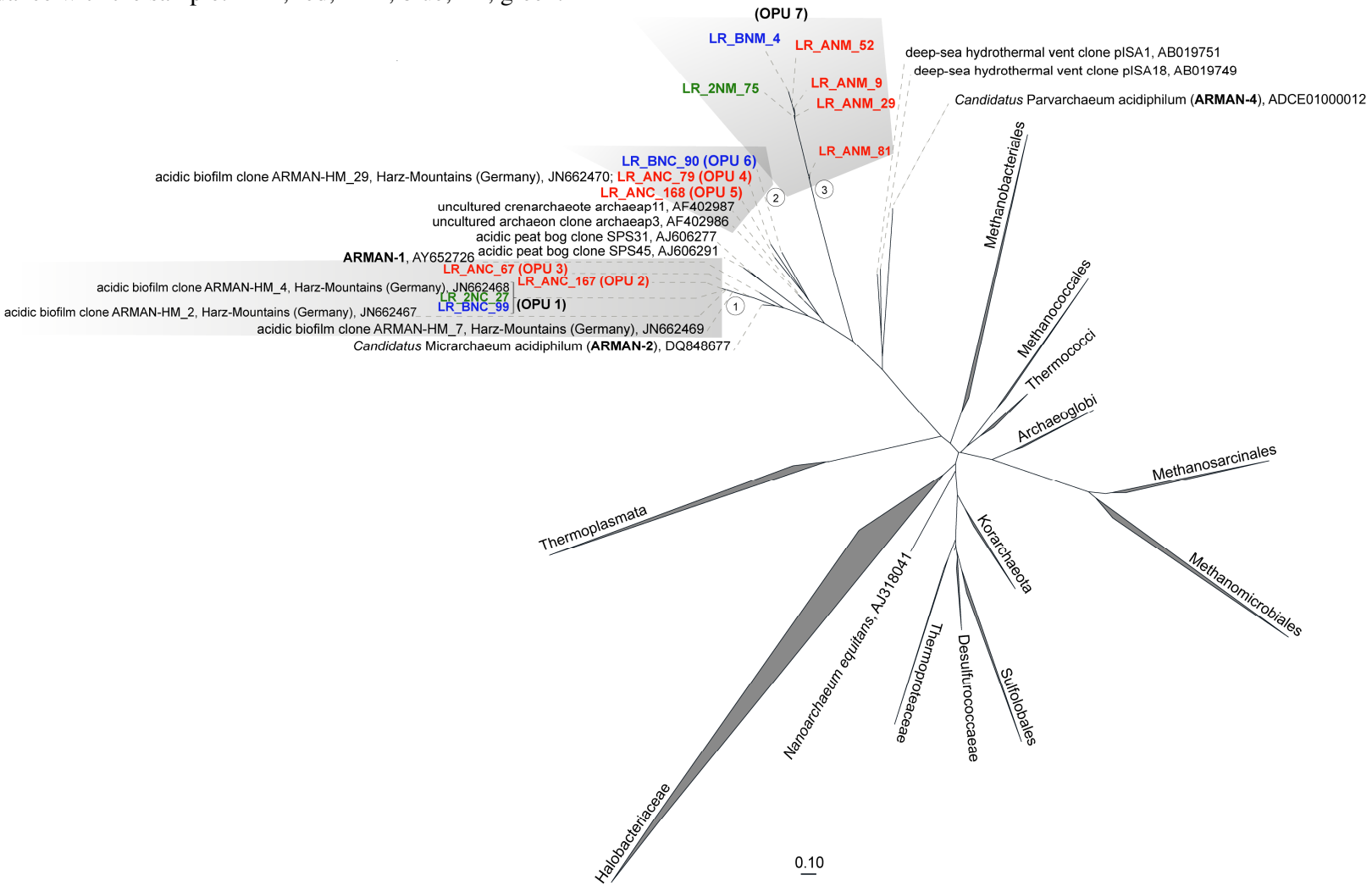


Fig. S5 Fluorescence microscopy images the main bacterial morphologies present during the colonization assay. DAPI stained cells (left) and bacterial probe (right) hybridizations at 6 h **(A)**, 2 days **(B)**, 10 days **(C),** 23 days, **(D)** and 58 days **(E)**.

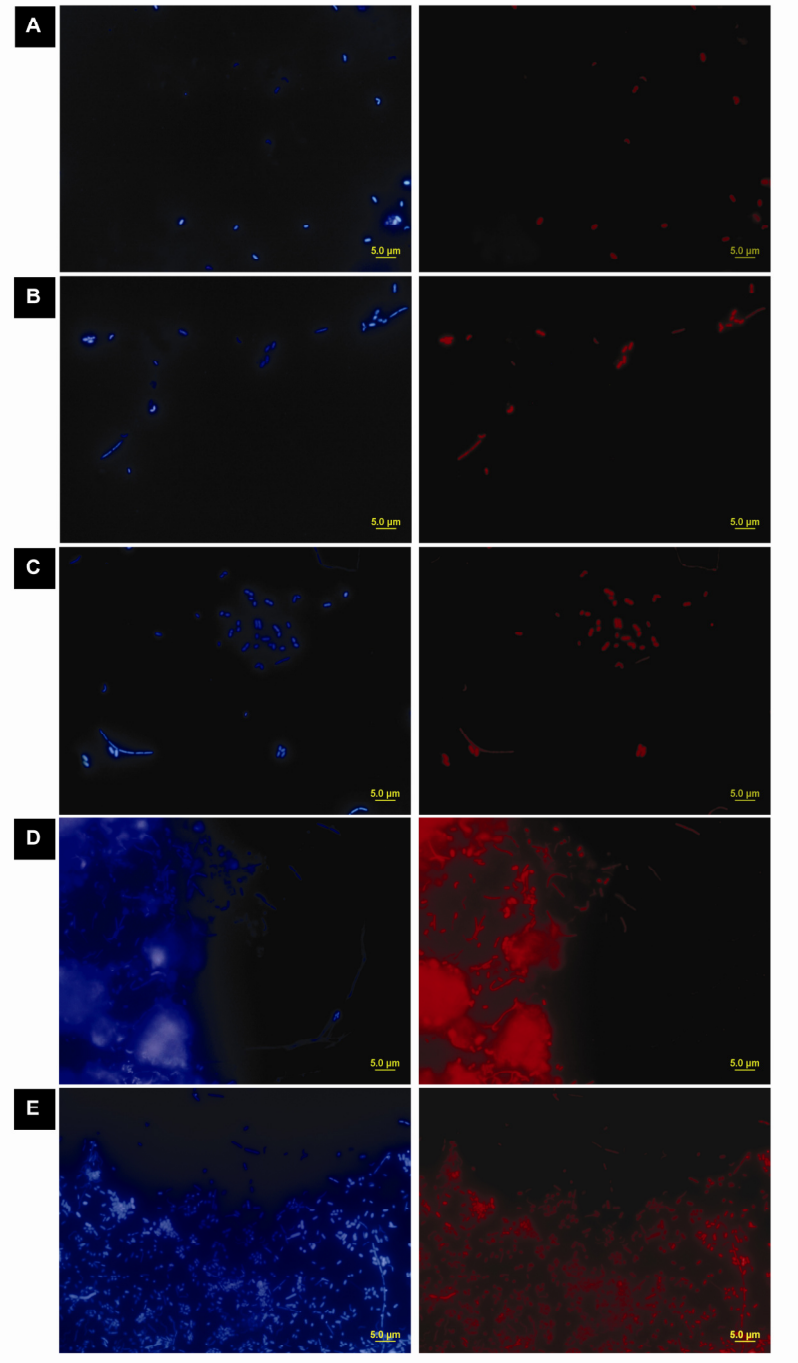


Fig. S6 Overview of the relative abundance of phylogenetic microbial groups implicated in the settlement of the initial B1 community. **(A)** Phylum, class and order-level taxonomic affiliations for the slides recovered bacterial sequences through SSU rRNA hypervariable tag pyro-sequencing analysis in comparison to the corresponding taxonomic profiles obtained for the mature B1 (A and B) and B2 communities. **(B)** Photography of the region where slides were submerged, where it becomes visible the discontinuity of B1 with the slight augmentation of the depth (inset); below, schematic representation of the biofilm-collector device used. **(C)** Temporal evolution in the proportion of the phylum (1), class (2) and order-level (3) taxonomic groups detected at all sampling times. Corresponding colored lines´ assignations are represented next to the taxa indicated in the (A) legends. **(D)** Rarefaction plot corresponding to the analyzed samples.

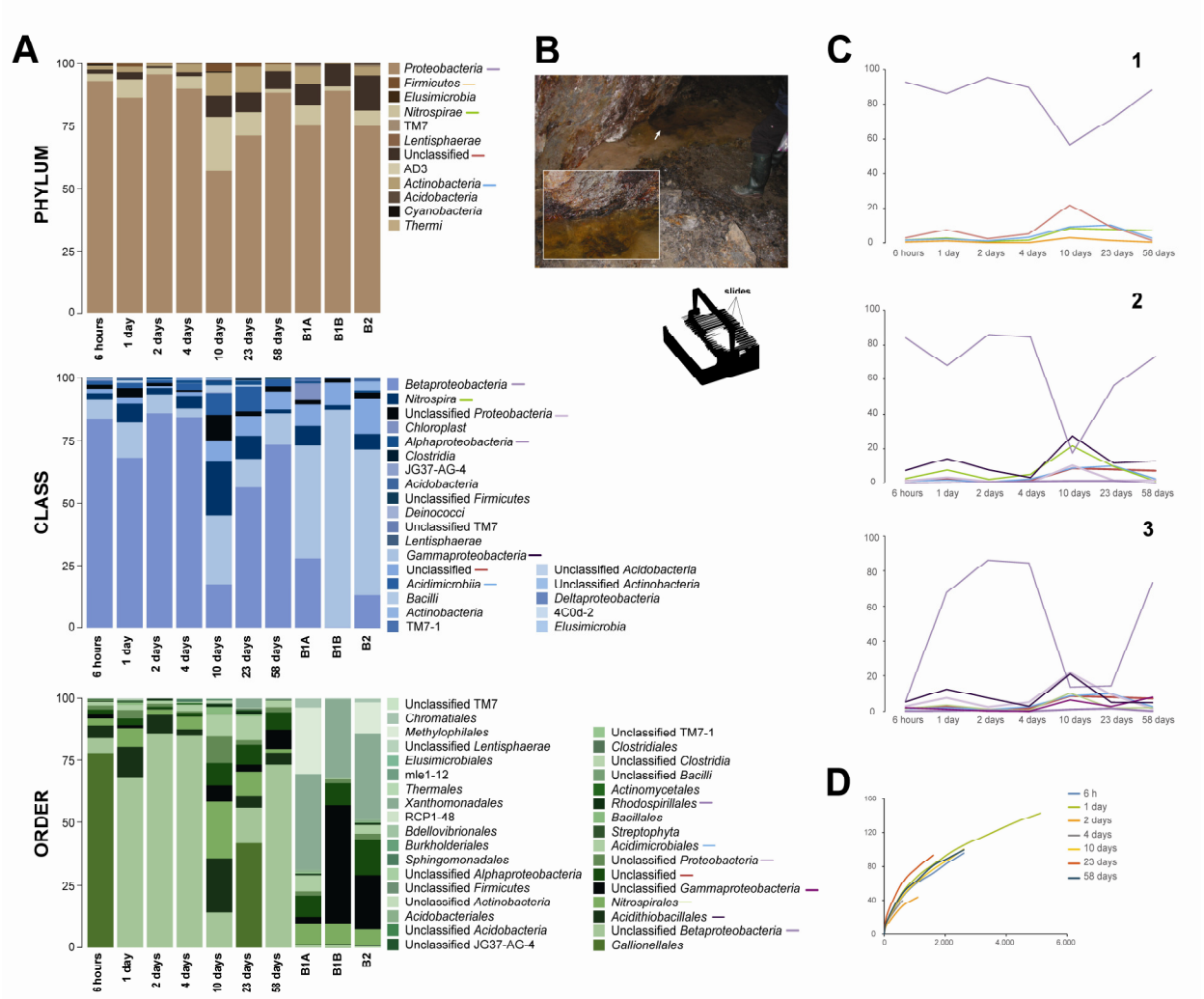


Fig. S7 X-ray photoelectron spectroscopy spectra for the Fe $2p_{3/2}$ and S 2p components over dried and homogenized B1A, B1B and B2 structures. Asterisk (*), arbitrary units.

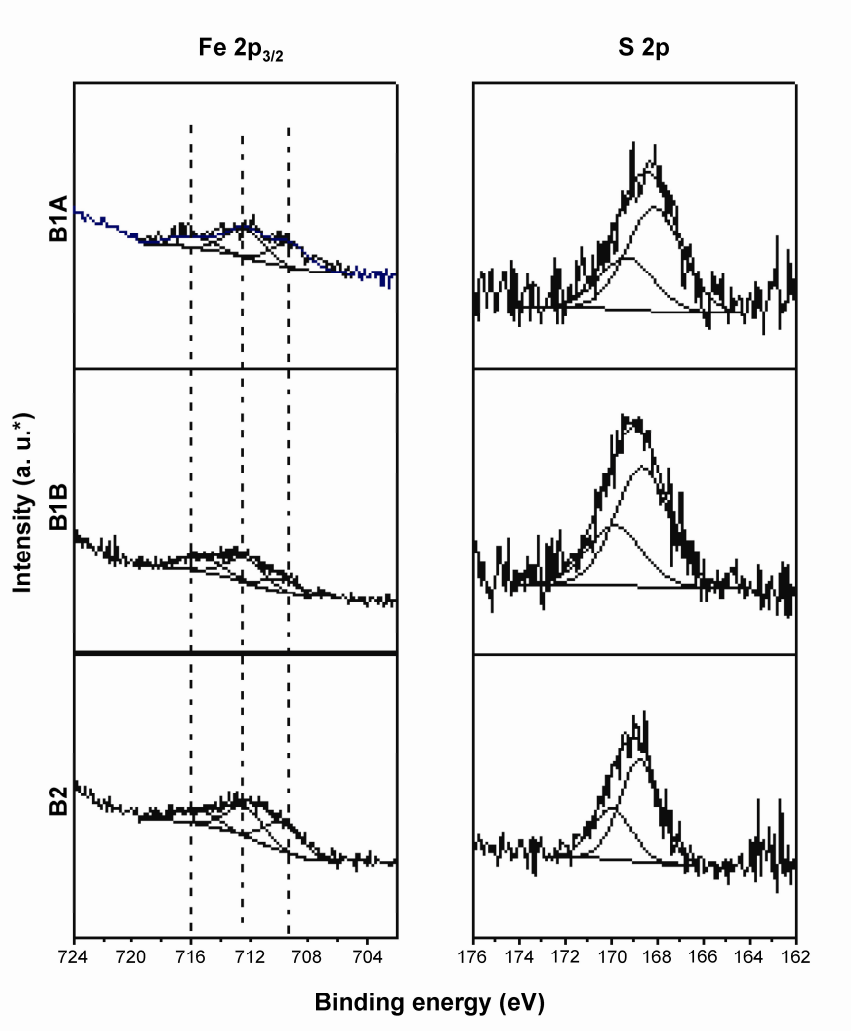


Fig. S8 Expression patterns and distributions of taxonomic categories of proteins that were identified in the B1, B1B and B2 metaproteomes. Expression pattern refers to the absolute protein abundance (mol protein/mol total protein) in each of the samples.

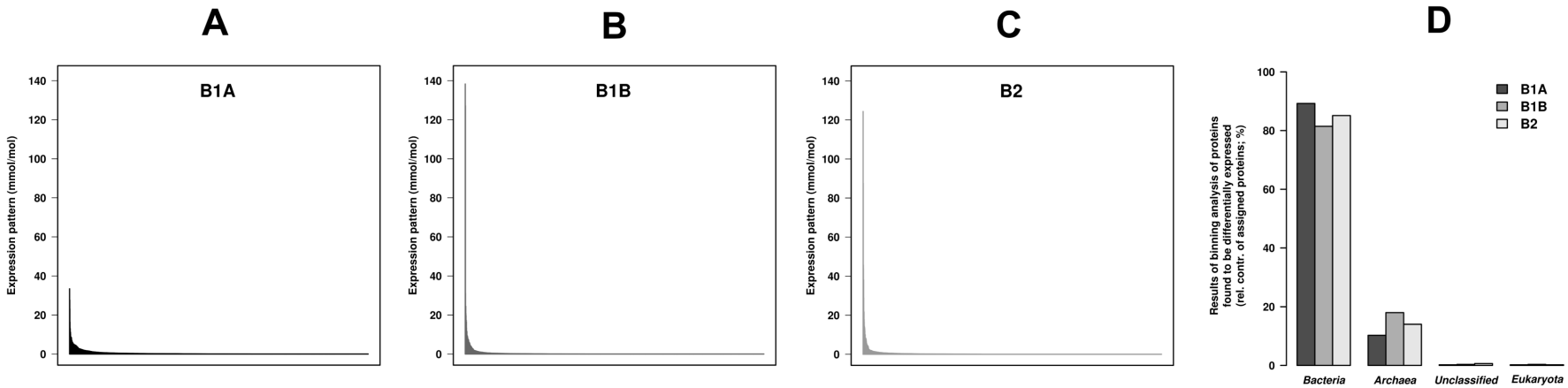


Fig. S9 Community behavior characterized by mapping to metabolic pathways. Proteome-scale metabolic reconstructions were based on the expression values of a set of proteins with Enzyme Commission (EC) numbers that were found among the 1,589 proteins that were unambiguously quantified in the metaproteomes. Ad-hoc computer programs written in C and Perl were used to obtain information from local copies of KEGG and to generate the graphical representations. Panels (**A**), (**B**) and (**C**) represent B1A, B1B and B2 community profiles, respectively. This Figure (as three independent panels) is provided as separate file (scalable PDF version) to ensure high quality resolution.

Panel A

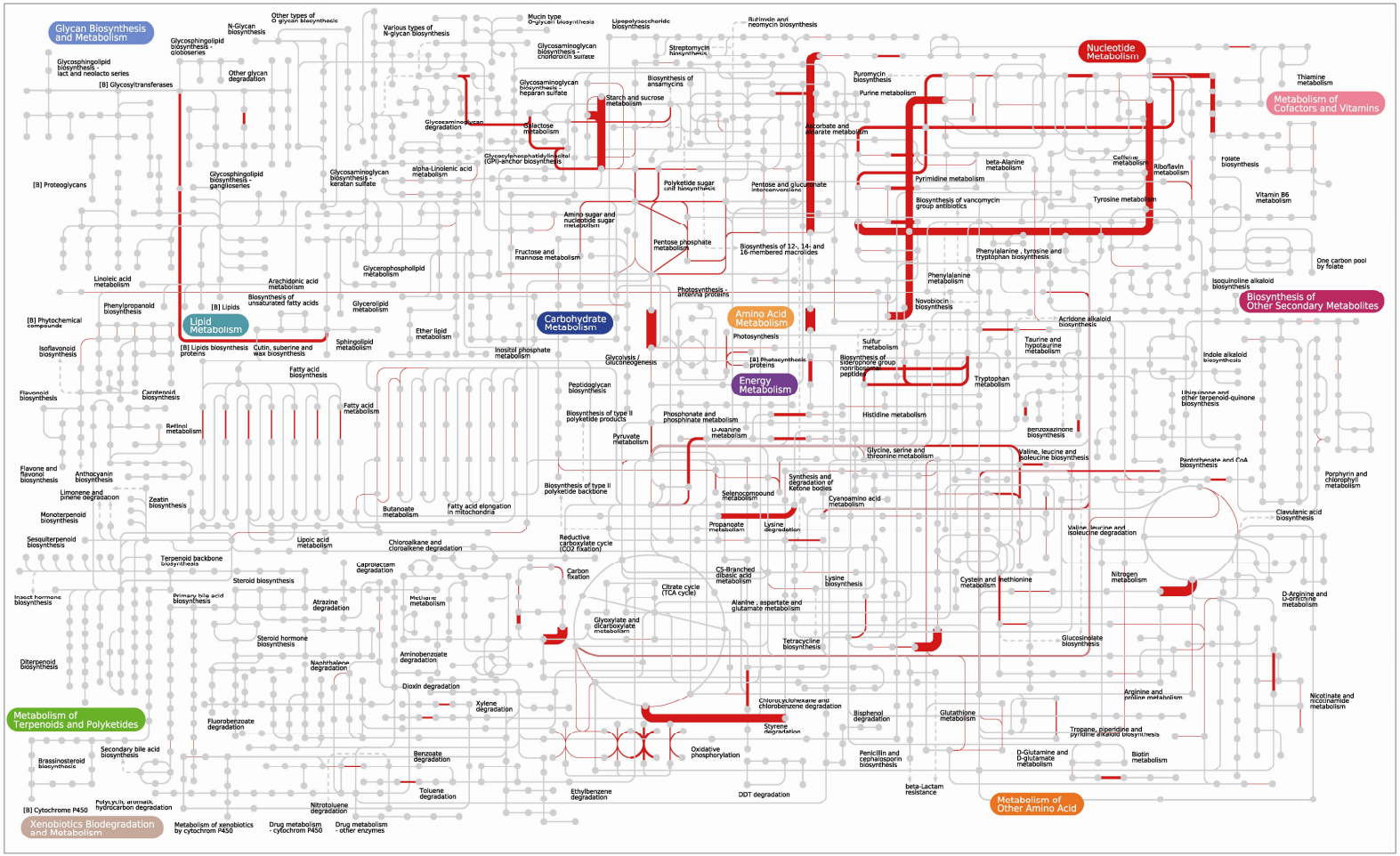


Fig. S9 Community behavior characterized by mapping to metabolic pathways. Proteome-scale metabolic reconstructions were based on the expression values of a set of proteins with Enzyme Commission (EC) numbers that were found among the 1,589 proteins that were unambiguously quantified in the metaproteomes. Ad-hoc computer programs written in C and Perl were used to obtain information from local copies of KEGG and to generate the graphical representations. Panels (**A**), (**B**) and (**C**) represent B1A, B1B and B2 community profiles, respectively. This Figure (as three independent panels) is provided as separate file (scalable PDF version) to ensure high quality resolution.
Panel B

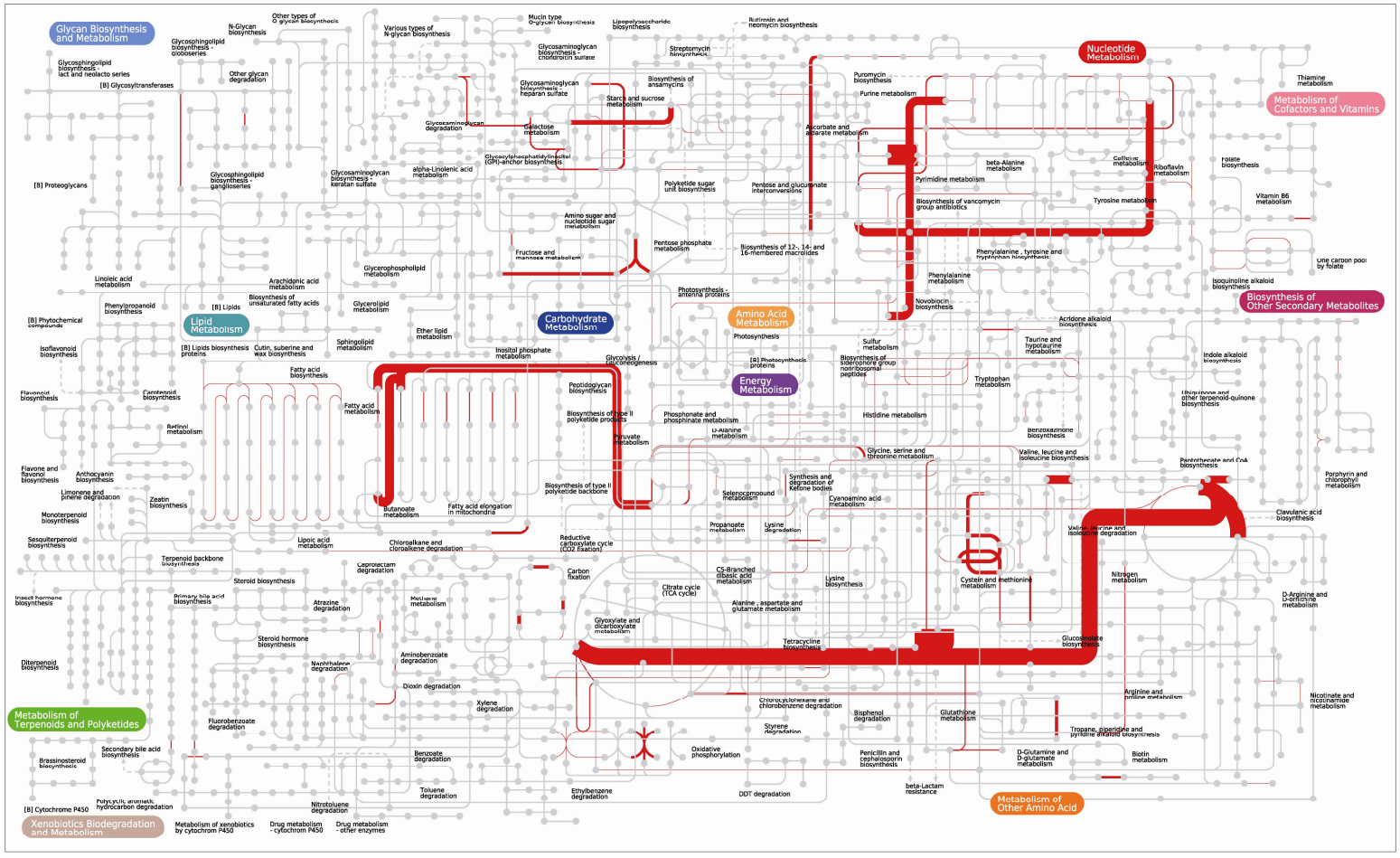


Fig. S9 Community behavior characterized by mapping to metabolic pathways. Proteome-scale metabolic reconstructions were based on the expression values of a set of proteins with Enzyme Commission (EC) numbers that were found among the 1,589 proteins that were unambiguously quantified in the metaproteomes. Ad-hoc computer programs written in C and Perl were used to obtain information from local copies of KEGG and to generate the graphical representations. Panels (**A**), (**B**) and (**C**) represent B1A, B1B and B2 community profiles, respectively. This Figure (as three independent panels) is provided as separate file (scalable PDF version) to ensure high quality resolution.
Panel C

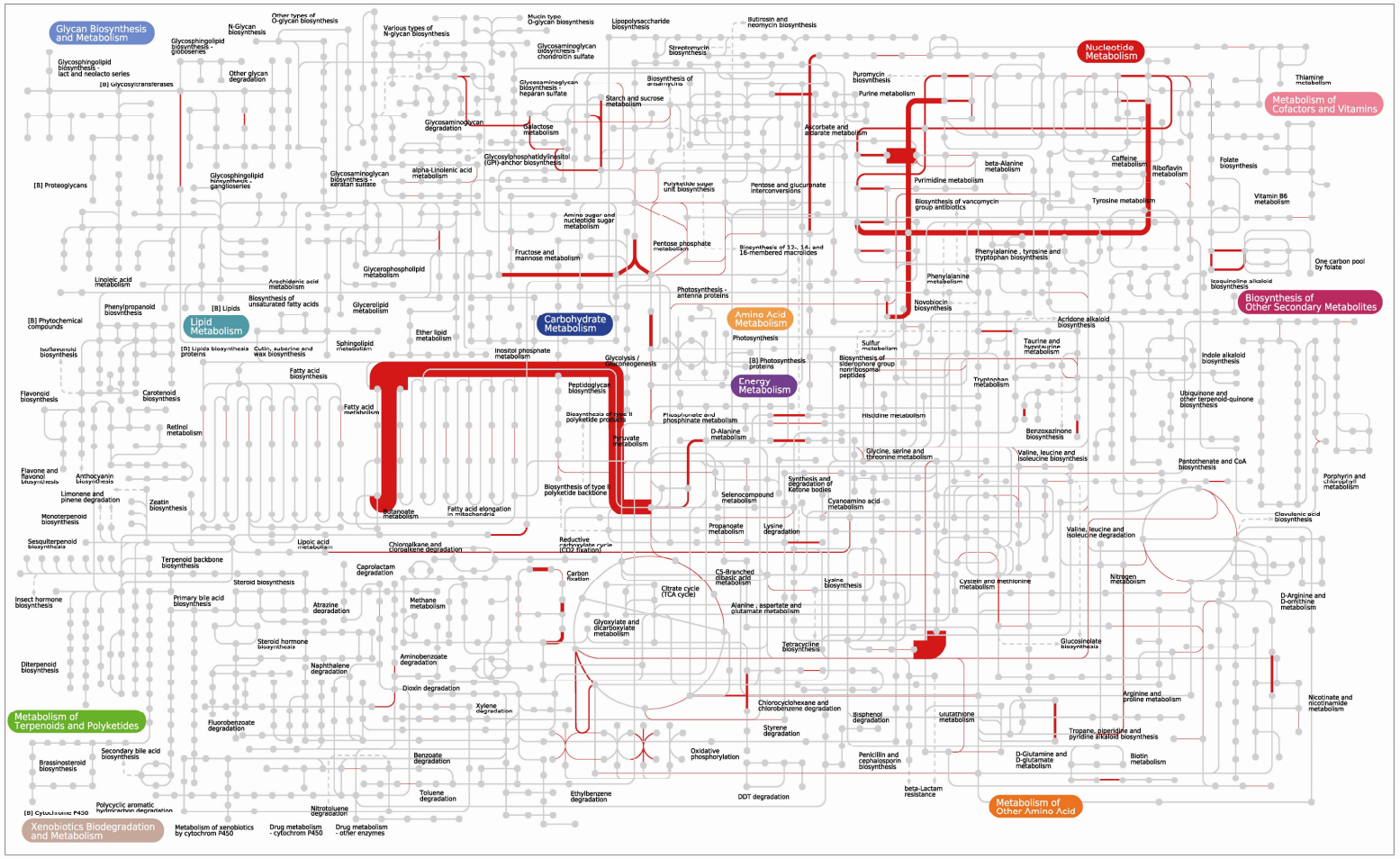


Fig. S10 Heat map and hierarchical cluster of the samples based on the relative contribution of each metabolic pathway. The contribution to each pathway was computed based on the average relative weight of the reactions linked to a given pathway present in each sample. KEGG identifiers used for the pathways are specifically shown. Codes as follows: ec00010 Glycolysis / Gluconeogenesis; ec00020 Citrate cycle (TCA cycle); ec00030 Pentose phosphate pathway; ec00040 Pentose and glucuronate interconversions; ec00051 Fructose and mannose metabolism; ec00052 Galactose metabolism; ec00053 Ascorbate and aldarate metabolism; ec00061 Fatty acid biosynthesis; ec00062 Fatty acid elongation; ec00071 Fatty acid metabolism; ec00072 Synthesis and degradation of ketone bodies; ec00120 Primary bile acid biosynthesis; ec00130 Ubiquinone and other terpenoid-quinone biosynthesis; ec00140 Steroid hormone biosynthesis; ec00190 Oxidative phosphorylation; ec00195 Photosynthesis; ec00230 Purine metabolism; ec00240 Pyrimidine metabolism; ec00250 Alanine, aspartate and glutamate metabolism; ec00260 Glycine, serine and threonine metabolism; ec00270 Cysteine and methionine metabolism; ec00280 Valine, leucine and isoleucine degradation ; ec00281 Geraniol degradation; ec00290 Valine, leucine and isoleucine biosynthesis; ec00300 Lysine biosynthesis; ec00310 Lysine degradation; ec00311 Penicillin and cephalosporin biosynthesis; ec00312 beta-Lactam resistance; ec00330 Arginine and proline metabolism; ec00340 Histidine metabolism; ec00350 Tyrosine metabolism; ec00360 Phenylalanine metabolism; ec00361 Chlorocyclohexane and chlorobenzene degradation; ec00362 Benzoate degradation; ec00364 Fluorobenzoate degradation; ec00380 Tryptophan metabolism; ec00400 Phenylalanine, tyrosine and tryptophan biosynthesis; ec00410 beta-Alanine metabolism; ec00430 Taurine and hypotaurine metabolism; ec00450 Selenocompound metabolism; ec00460 Cyanoamino acid metabolism; ec00471 D-Glutamine and D-glutamate metabolism; ec00473 D-Alanine metabolism; ec00480 Glutathione metabolism; ec00500 Starch and sucrose metabolism; ec00511 Other glycan degradation; ec00520 Amino sugar and nucleotide sugar metabolism; ec00521 Streptomycin biosynthesis; ec00523 Polyketide sugar unit biosynthesis; ec00524 Butirosin and neomycin biosynthesis; ec00531 Glycosaminoglycan degradation; ec00540 Lipopolysaccharide biosynthesis; ec00550 Peptidoglycan biosynthesis; ec00561 Glycerolipid metabolism; ec00564 Glycerophospholipid metabolism; ec00592 alpha-Linolenic acid metabolism; ec00600 Sphingolipid metabolism; ec00604 Glycosphingolipid biosynthesis - ganglio series; ec00620 Pyruvate metabolism; ec00622 Xylene degradation; ec00623 Toluene degradation; ec00625 Chloroalkane and chloroalkene degradation; ec00626 Naphthalene degradation; ec00627 Aminobenzoate degradation; ec00630 Glyoxylate and dicarboxylate metabolism; ec00633 Nitrotoluene degradation; ec00640 Propanoate metabolism; ec00643 Styrene degradation; ec00650 Butanoate metabolism; ec00660 C5-Branched dibasic acid metabolism; ec00670 One carbon pool by folate; ec00680 Methane metabolism; ec00710 Carbon fixation in photosynthetic organisms; ec00720 Carbon fixation pathways in prokaryotes; ec00730 Thiamine metabolism; ec00740 Riboflavin metabolism; ec00750 Vitamin B6 metabolism; ec00760 Nicotinate and nicotinamide metabolism; ec00770 Pantothenate and CoA biosynthesis; ec00780 Biotin metabolism; ec00790 Folate biosynthesis; ec00830 Retinol metabolism; ec00860 Porphyrin and chlorophyll metabolism; ec00900 Terpenoid backbone biosynthesis; ec00903 Limonene and pinene degradation; ec00910 Nitrogen metabolism; ec00920 Sulfur metabolism; ec00930 Caprolactam degradation; ec00940 Phenylpropanoid biosynthesis; ec00941 Flavonoid biosynthesis; ec00966 Glucosinolate biosynthesis; ec00980 Metabolism of xenobiotics by cytochrome P450; ec00982 Drug metabolism - cytochrome P450; ec00983 Drug metabolism - other enzymes; ec01040 Biosynthesis of

unsaturated fatty acids; ec01051 Biosynthesis of ansamycins. The colour scale (defined as colour key in the figure) quantifies the relative percentages from one sample to the other, and goes from red (highly abundant) to black (abundant) to light green (low abundant).

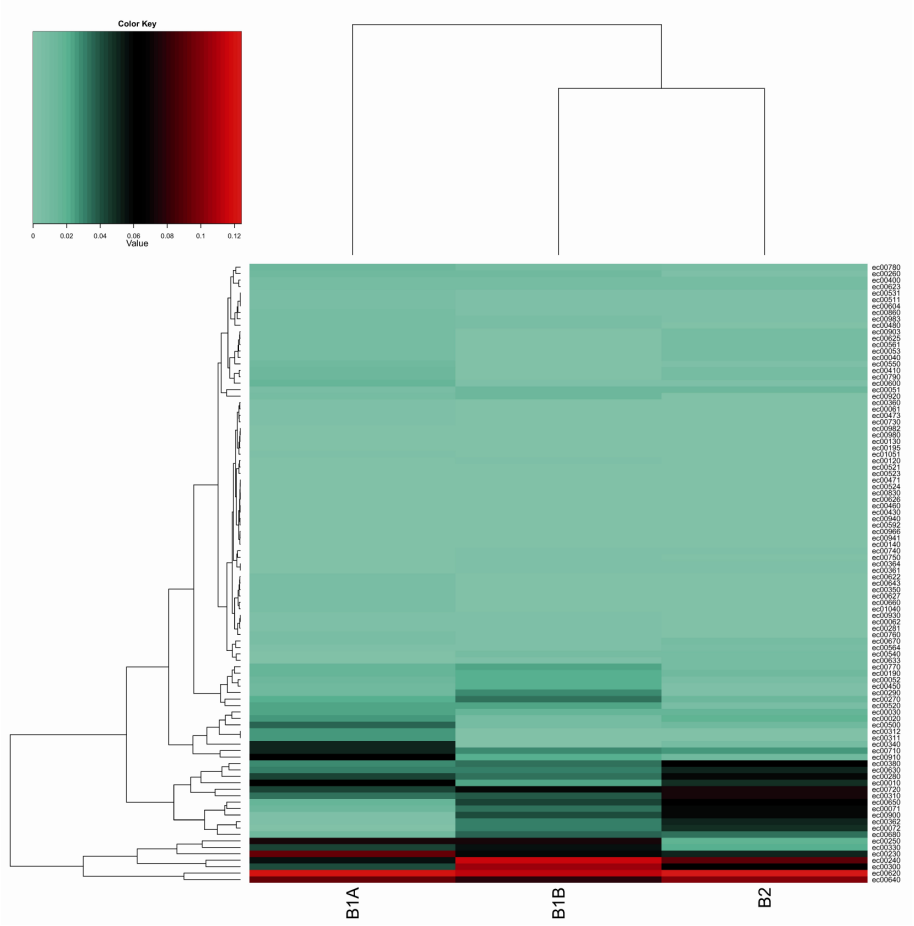


Fig. S11. Shifts in carbohydrate turnover profile in the B1 streamer. Enzyme activities (units/g total protein) from the samples obtained from the B1 streamer at different depths against 14 different sugar substrates (as *p*NP derivatives) were quantified by measuring the release of *p*NP in triplicates. A single plot, with mean values estimated for each group of samples using three independent measurements each (SD lower than 5%) is shown.

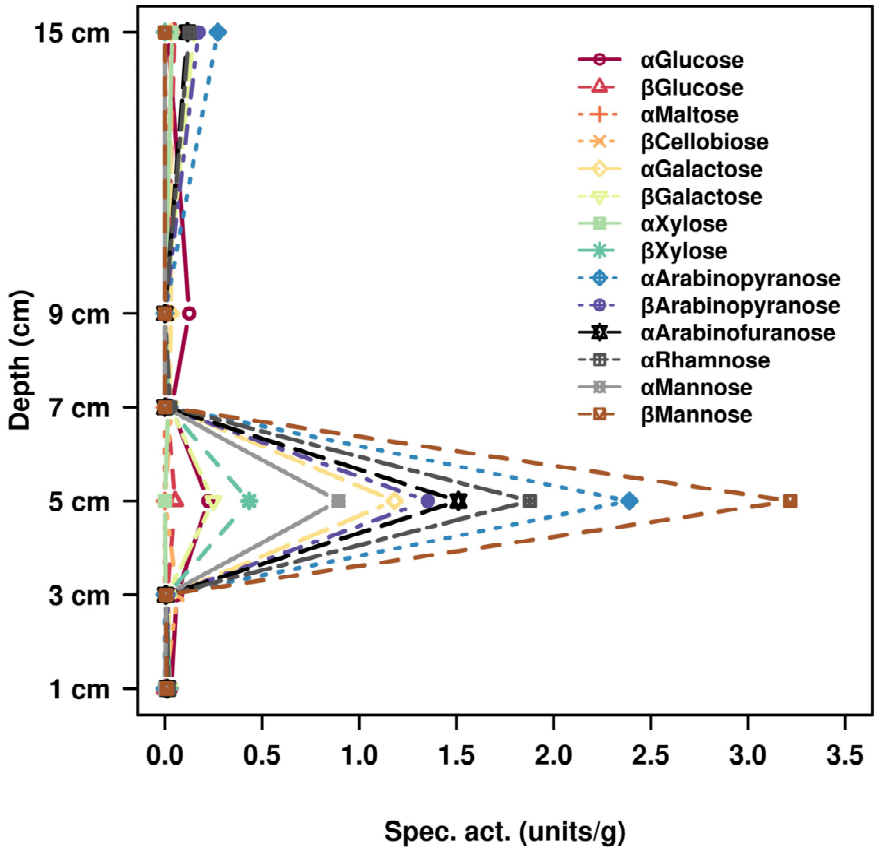
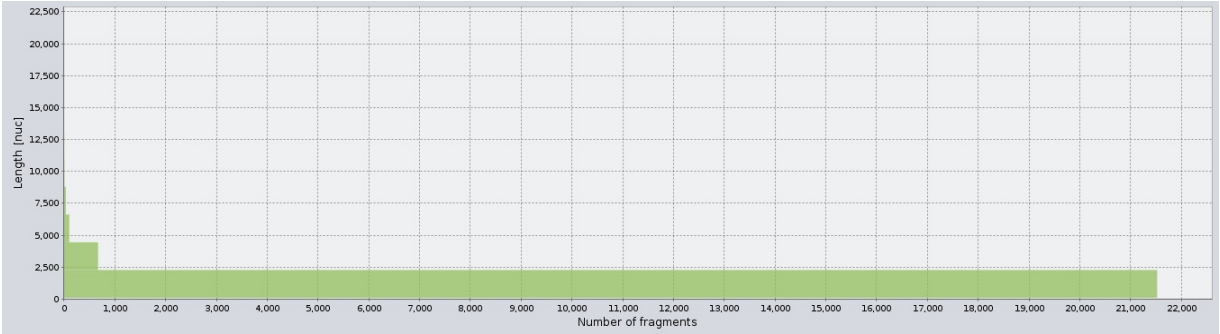
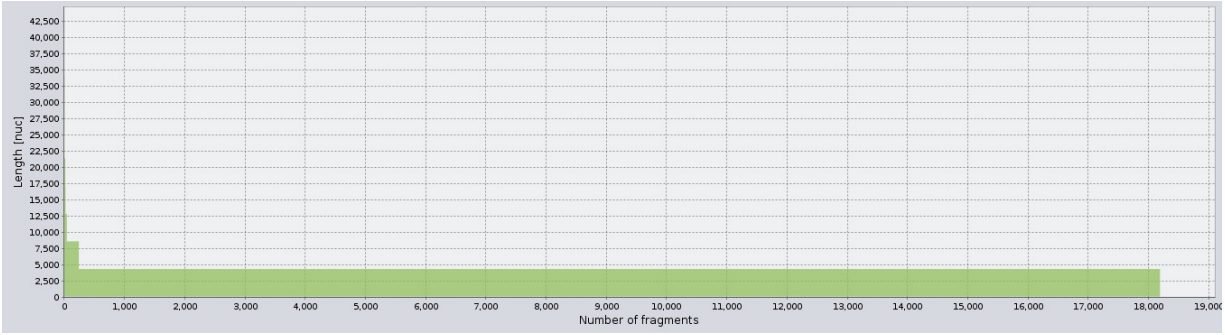


Fig. S12. Contigs length distribution of B1A (A), B1B (B) and B2 (C) metagenomes.

(A)



(B)



(C)